J. Math. Anal. Appl. 382 (2011) 822–842

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Recursive estimation for ordered eigenvectors of symmetric matrix with observation noise ☆

Han-Fu Chen *, Hai-Tao Fang, Li-Li Zhang

Key Laboratory of Systems and Control, Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing 100080, PR China

ARTICLE INFO

Article history:
Received 5 July 2010
Available online 30 April 2011
Submitted by Y. Huang

Keywords:
Principal component analysis (PCA)
Recursive algorithm
Stochastic approximation
Ordered convergence
Convergence rate

ABSTRACT

The principal component analysis is to recursively estimate the eigenvectors and the corresponding eigenvalues of a symmetric matrix A based on its noisy observations $A_k = A + N_k$, where A is allowed to have arbitrary eigenvalues with multiplicity possibly bigger than one. In the paper the recursive algorithms are proposed and their ordered convergence is established: It is shown that the first algorithm a.s. converges to a unit eigenvector corresponding to the largest eigenvalue, the second algorithm a.s. converges to a unit eigenvector corresponding to either the second largest eigenvalue in the case the largest eigenvalue is of single multiplicity or the largest eigenvalue if the multiplicity of the largest eigenvalue is bigger than one, and so on. The convergence rate is also derived.

1. Introduction

With a practical system there may be a large amount of variables involved, and each variable may provide a certain information. However, problems connected with a system with a huge number of variables are difficult to deal with. It is often the case that the variables are correlated and they are not equally important. The PCA proposed by Pearson [9] aims at treating problems involving a large amount of variables, and its idea is to estimate eigenvectors corresponding to eigenvalues stated in the non-increasing order, i.e., first to select the most important factors and then the less important factors in the decreasing order of importance by using some linear transformations acting on the variables.

PCA is now widely used in various areas such as data analysis, image compression, pattern recognition, subspace identification, and many others, see, e.g., [4,6,7] among others.

Let each component of $x \in \mathbb{R}^n$ represent a variable of the system with huge n. PCA is to find the eigenvectors of $A \triangleq Exx^T$. In this setting, $x_k x_k^T$ serves as an observation on A.

We now consider a slightly modified setting. Let a deterministic symmetric matrix A with arbitrary eigenvalues be noisily observed: $A_k = A + N_k$, where A_k is the observation at time k and N_k is the observation noise. On the basis of observations $\{A_k\}$, it is required to recursively estimate the eigenvectors and the corresponding eigenvalues of A in the decreasing order of eigenvalues.

For the case where all eigenvalues of A are of single multiplicity the ordered convergence of a stochastic approximation (SA) type algorithm has been established in [7,8] under rather restrictive constraints, while in [11] the convergence has been proved under much weaker conditions but the required ordering is not guaranteed.

☆ The work is supported by NSFC under Grants 60821091 and 60874001 for Han-Fu Chen and Li-Li Zhang, and under Grants 60221301 and 60774020 for Hai Tao Fang, and by a grant from the National Laboratory of Space Intelligent Control for all of the authors.

* Corresponding author.
E-mail address: hfchen@iss.ac.cn (H.-F. Chen).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2011.04.072

The case where the eigenvalues of A may have multiplicity bigger than one is dealt in [1,5,10]. The ordered convergence is established in [1] under some restrictive conditions including that all eigenvalues should be positive, while in [10] the ordered convergence is not achieved. The limiting behavior of the algorithm is not completely clarified in [5].

In this paper we define the PCA algorithms in Section 2, and the distance between the estimate given by the algorithms and the subspace of the corresponding unit eigenvectors is proved to converge to zero in Section 3. In Section 4 it is shown that the convergence is actually ordered in accordance with eigenvalues stated in a non-increasing order. In Section 5 it is shown that each estimate converges to a unit eigenvector, and the convergence rate is pointed out as well. A few concluding remarks are given in Section 6. Some auxiliary results are presented in Appendix A.

2. Algorithms

Let us first define the recursive algorithm for $\{u_k^{(1)}\}_{k\geqslant 0}$ estimating the normalized eigenvector or one of the normalized eigenvectors corresponding to the largest eigenvalue of A:

$$\tilde{u}_{k+1}^{(1)} = u_k^{(1)} + a_k A_{k+1} u_k^{(1)}, \quad a_k > 0, \tag{1}$$

$$u_{k+1}^{(1)} = \tilde{u}_{k+1}^{(1)} / \|\tilde{u}_{k+1}^{(1)}\|, \tag{2}$$

whenever $\|\tilde{u}_{k+1}^{(1)}\| \neq 0$.

In the case where $\|\tilde{u}_{k+1}^{(1)}\| = 0$, $u_k^{(1)}$ is reset to be some other unit vector making $\|\tilde{u}_{k+1}^{(1)}\| \neq 0$ defined by (1).

Assuming $u_s^{(i)}$, $i = 1, \dots, j$, $s = 0, 1, \dots, k$ have been defined, we define $u_s^{(j+1)}$, $s = 0, 1, \dots, k+1$ as follows.

For this we first define the $n \times j$-matrix

$$V_s^{(j)} \triangleq \left[u_s^{(1)}, P_s^{(1)} u_s^{(2)}, \dots, P_s^{(j-1)} u_s^{(j)}\right], \qquad P_s^{(0)} \triangleq I, \quad s = 1, \dots, k, \tag{3}$$

where $P_s^{(i)} \triangleq I - V_s^{(i)} V_s^{(i)+}$, $i = 1, \dots, j-1$ with $V_s^{(i)+}$ being the pseudo-inverse of $V_s^{(i)}$ is the projection to the subspace orthogonal to the space spanned by columns of $V_s^{(i)}$.

Given an initial unit vector $u_0^{(j+1)}$, recursively define

$$\tilde{u}_{k+1}^{(j+1)} = P_k^{(j)} u_k^{(j+1)} + a_k P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}, \tag{4}$$

$$u_{k+1}^{(j+1)} = \tilde{u}_{k+1}^{(j+1)} / \|\tilde{u}_{k+1}^{(j+1)}\|, \tag{5}$$

whenever $\|\tilde{u}_{k+1}^{(j+1)}\| \neq 0$.

Otherwise, reset $u_k^{(j+1)}$ to be a unit vector so that $\|u_k^{(j+1)}\| = 1$ and $\|P_k^{(j)} u_k^{(j+1)}\| = 1$.

Noticing

$$a_k P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} \to 0$$

by A1 and A2 to be given later and $\|P_k^{(j)} u_k^{(j+1)}\| = 1$ after a resetting, we find that $\|\tilde{u}_{k+1}^{(j+1)}\| = 0$ may occur at most a finite number of times, and hence resetting $u_k^{(j+1)}$ ceases in a finite number of steps.

From now on we always assume that k is large enough and no resetting occurs.

For eigenvalues of A the recursive estimates $\{\lambda_k^{(j)}\}_{k\geqslant 1}$, $j = 1, \dots, n$ with arbitrary initial values $\lambda_0^{(j)}$ are given by the following algorithms

$$\lambda_{k+1}^{(j)} = \lambda_k^{(j)} - a_k\left(\lambda_k^{(j)} - u_k^{(j)T} A_{k+1} u_k^{(j)}\right). \tag{6}$$

Denote by J the set of all unit eigenvectors of A.

Let $V(J) \triangleq \{\lambda^{(1)}, \dots, \lambda^{(n)}\}$ be the set of eigenvalues of A stated in the non-increasing order. Notice that some of eigenvalues may be identical. The convergence analysis is completed by three steps.

Step 1. We first show that for each $\{u_k^{(j)}\}$ there exists a subset J_j of J such that as k tends to infinity $d(u_k^{(j)}, J_j)$, the distance between $u_k^{(j)}$ and J_j, converges to zero, and $\lambda_k^{(j)}$ converges to the eigenvalue $\lambda(j)$ associated with J_j.

Step 2. It is shown that the convergence established in Step 1 is ordered in the sense that $\lambda(j) = \lambda^{(j)}$. In other words, J_1 corresponds to the largest eigenvalue of A, and J_2 either coincides with J_1 in the case $\lambda^{(1)}$ is with multiplicity greater than one, or corresponds to the second largest eigenvalue of A and so on.

Step 3. Except the case where all eigenvalues are equal, it is shown there is a unit vector $u^{(j)} \in J_j$ such that $u_k^{(j)} \xrightarrow[k\to\infty]{} u^{(j)}$ and $\|u_k^{(j)} - u^{(j)}\| = O(a_k^{\delta})$ with $\delta > 0$.

For establishing results stated in Step 1, the following assumptions A1 and A2 are needed.

A1. $a_k > 0$, $a_k \xrightarrow[k\to\infty]{} 0$, $\sum_{k=0}^{\infty} a_k = \infty$, and $\sum a_k^{1+\eta} < \infty$ for any $\eta > 0$. Moreover, there is an $a > 0$ such that

$$\lim_{k\to\infty} a_{k+1}^{-1} - a_k^{-1} = a > 0.$$

A2. $A_k = A + N_k$, $\{N_k, \mathcal{F}_k\}$ is a bounded martingale difference sequence (mds) with $\sup_k \|N_{k+1}\| = \zeta < \infty$ a.s., and

$$\sup_k E\big(\|N_{k+1}\|^2 \bigm| \mathcal{F}_k\big) < \infty \quad \text{a.s.},$$

where $\mathcal{F}_k$ is the σ-algebra generated by $\{N_1, \ldots, N_k\}$.

Remark 1. It is clear that if $a_k = \frac{c}{k}$ with $c > 0$, then $\lim_{k\to\infty} a_{k+1}^{-1} - a_k^{-1} = \frac{1}{c}$, and A1 holds. Under A1, $\frac{a_k}{a_{k+1}} = 1 + O(a_k)$. So, in the sequel, we will not distinguish between $O(a_k)$ and $O(a_{k-s})$ for any finite s.

3. Convergence of estimates

In this section we show the convergence of (1)–(6).

Theorem 1. *Assume* A1 *and* A2 *hold. Then estimates* $\{u_k^{(i)}\}$, $i = 1, \ldots, n$ *given by* (1)–(5) *have the following properties*:

(i) *There exists a connected subset* J_i *of* J *such that* $d(u_k^{(i)}, J_i) \xrightarrow[k\to\infty]{} 0$.

(ii) *There is an eigenvalue* $\lambda(i) \in V(J)$ *so that*

$$d\big(Au_k^{(i)}, \lambda(i)u_k^{(i)}\big) \to 0, \quad \textit{and} \quad Au = \lambda(i)u \quad \textit{for any } u \in J_i. \tag{7}$$

(iii) *The recursive algorithm for* $u_k^{(i)}$ *can be expressed as*

$$u_{k+1}^{(i)} = u_k^{(i)} + a_k\big(Au_k^{(i)} - \big(u_k^{(i)T} A u_k^{(i)}\big)u_k^{(i)}\big) + O\left(a_k\left(a_k + \sum_{s=1}^{i-1} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k\delta_{k+1}(i), \tag{8}$$

where $\delta_{k+1}(i)$ *is bounded and is a linear combination of martingale difference sequences* (*mds*) *being measurable with respect to* $\mathcal{F}_{k+1}, \mathcal{F}_k, \ldots,$ *and* $\mathcal{F}_{k+2-i}$, *respectively, and with bounded second conditional moments and* $\sum_{k=1}^{\infty} a_k\delta_{k+1}(i) < \infty$ *a.s.*

(iv) $\lambda_k^{(j)}$ *defined by* (6) *converges to* $\lambda(j)$ *as* k *tends to infinity,* $j = 1, \ldots, n$.

Proof. For large k, $u_{k+1}^{(1)}$ can be expanded as follows

$$\begin{aligned}
u_{k+1}^{(1)} &= \big(u_k^{(1)} + a_k A_{k+1} u_k^{(1)}\big)\big(1 + 2a_k u_k^{(1)T} A_{k+1} u_k^{(1)} + a_k^2 u_k^{(1)T} A_{k+1}^2 u_k^{(1)}\big)^{-\frac{1}{2}} \\
&= \big(u_k^{(1)} + a_k A_{k+1} u_k^{(1)}\big)\big\{1 - a_k u_k^{(1)T} A_{k+1} u_k^{(1)} + O\big(a_k^2\big)\big\} \\
&= u_k^{(1)} + a_k A_{k+1} u_k^{(1)} - a_k\big(u_k^{(1)T} A_{k+1} u_k^{(1)}\big)u_k^{(1)} + O\big(a_k^2\big) \\
&= u_k^{(1)} + a_k\big(Au_k^{(1)} - \big(u_k^{(1)T} A u_k^{(1)}\big)u_k^{(1)}\big) + a_k\varepsilon_{k+1}^{(1)} + O\big(a_k^2\big),
\end{aligned} \tag{9}$$

where $\varepsilon_{k+1}^{(1)} = N_{k+1}u_k^{(1)} - (u_k^{(1)T} N_{k+1} u_k^{(1)})u_k^{(1)}$.

By A2, $(\varepsilon_{k+1}^{(1)}, \mathcal{F}_{k+1})$ is an mds. Since $\sum_k a_k^2 < \infty$, $\sup_k E\{\|N_{k+1}\|^2 \mid \mathcal{F}_k\} < \infty$, $\|u_k^{(1)}\| = 1$, by the convergence theorem for mds we have

$$\sum_k a_k\big[N_{k+1}u_k^{(1)} - \big(u_k^{(1)T} N_{k+1} u_k^{(1)}\big)u_k^{(1)}\big] < \infty \quad \text{a.s.} \tag{10}$$

Thus, (8) with $\delta_{k+1}(1) = \varepsilon_{k+1}^{(1)}$ holds for $i = 1$.

From (10) it follows that

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T}\left\| \sum_{i=k}^{m(k,T)} a_i\big[N_{i+1}u_i^{(1)} - \big(u_i^{(1)T} N_{i+1} u_i^{(1)}\big)u_i^{(1)}\big] \right\| = 0, \tag{11}$$

and hence

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\left\|\sum_{i=k}^{m(k,T)}a_i\big(\varepsilon_{i+1}^{(1)}+O(a_i)\big)\right\|=0, \tag{12}$$

where $m(k,T)\triangleq\max\{m:\ \sum_{i=k}^{m}a_i\leqslant T\}$.

Define $f(u)\triangleq Au-(u^TAu)u$ on the unit sphere S. It is clear that the root set of $f(\cdot)$ on S is J.

Let $v(u)\triangleq-\frac{1}{2}u^TAu$. Then, for $u\in S$ we have

$$\begin{aligned}v_u^T(u)f(u)&=-u^TA\big[Au-\big(u^TAu\big)u\big]=-u^TA^2u+\big(u^TAu\big)^2\\&\begin{cases}<\|Au\|^2\|u\|^2-u^TA^2u=0, & \text{if } u\notin J,\\=0, & \text{if } u\in J.\end{cases}\end{aligned} \tag{13}$$

Denote by J_1 the totality of the limiting points of $u_k^{(1)}$. By (12) applying Theorem 2.2.3 and Remark 2.2.6 of [3] to (9) leads to

$$d\big(u_k^{(1)T}Au_k^{(1)},V(J)\big)\to 0 \quad\text{and}\quad d\big(u_k^{(1)},J_1\big)\to 0, \tag{14}$$

where J_1 is a connected subset of J.

Since $V(J)$ is composed of isolated points, by (14) there is a $\lambda(1)\in V(J)$ such that

$$d\big(u_k^{(1)T}Au_k^{(1)},\lambda(1)\big)\to 0. \tag{15}$$

We now show that

$$Au=\lambda(1)u,\quad \forall u\in J_1. \tag{16}$$

Assume the converse: there exist $\tilde u\in J_1$ and $\lambda(1)'\neq\lambda(1)$ such that

$$A\tilde u=\lambda(1)'\tilde u \quad\text{and}\quad \tilde u^TA\tilde u=\lambda(1)'. \tag{17}$$

Since J_1 is composed of limiting points of $\{u_k^{(1)}\}$, for $\tilde u\in J_1$ there must exist a subsequence $\{u_{n_k}^{(1)}\}$ such that $u_{n_k}^{(1)}\to\tilde u$. By (17) it follows that

$$d\big(u_{n_k}^{(1)T}Au_{n_k}^{(1)},\lambda(1)'\big)\to 0,$$

which contradicts with (15). Hence, (16) holds.

Since $d(u_k^{(1)},J_1)\to 0$ by (14), from (16) it follows that $d(Au_k^{(1)},\lambda(1)u_k^{(1)})\xrightarrow[k\to\infty]{}0$.

Thus, we have proved the theorem for $i=1$.

We need to show that

$$\big(V_{k+1}^{(i-1)T}V_{k+1}^{(i-1)}\big)^{-1}=I+O\big(a_k^2\big), \tag{18}$$

$$V_{k+1}^{(i-1)T}u_{k+1}^{(i)}=O\left(a_k\left(a_k+\sum_{s=1}^{i-1}d\big(u_k^{(s)},J_s\big)\right)\right)+a_k\eta_{k+1}^{(i-1)}, \tag{19}$$

and

$$V_{k+1}^{(i-1)}V_{k+1}^{(i-1)+}u_{k+1}^{(i)}=O\left(a_k\left(a_k+\sum_{s=1}^{i-1}d\big(u_k^{(s)},J_s\big)\right)\right)+a_k\gamma_{k+1}(i) \tag{20}$$

are valid for all i: $2\leqslant i\leqslant n$, where both $\eta_{k+1}^{(i-1)}$ and $\gamma_{k+1}(i)$ are the linear combinations of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+3-i}$, respectively, and with bounded second conditional moments. Thus, $\sum_{k=1}^{\infty}a_k\eta_{k+1}^{(i)}<\infty$, and $\sum_{k=1}^{\infty}a_k\gamma_{k+1}(i)<\infty$.

Since $V_{k+1}^{(1)}=u_{k+1}^{(1)}$, from (2) it is seen that $u_{k+1}^{(1)T}u_{k+1}^{(1)}=1$, and hence (18) is valid for $i=2$.

Let us prove (19), (20), and the theorem for $i=2$.

We have

$$\begin{aligned}\|\tilde u_{k+1}^{(2)}\|^{-1}&=\big\{\big[P_k^{(1)}u_k^{(2)}+a_kP_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big]^T\big[P_k^{(1)}u_k^{(2)}+a_kP_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big]\big\}^{-\frac{1}{2}}\\&=\big[u_k^{(2)T}P_k^{(1)}u_k^{(2)}+2a_ku_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}+a_k^2u_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big]^{-\frac{1}{2}}\\&=\big[1-u_k^{(2)T}V_k^{(1)}V_k^{(1)+}u_k^{(2)}+2a_ku_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}+a_k^2u_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big]^{-\frac{1}{2}}\\&=1+\frac{1}{2}u_k^{(2)T}V_k^{(1)}V_k^{(1)T}u_k^{(2)}-a_ku_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}+O\big(a_k^2\big).\end{aligned} \tag{21}$$

By (9) and noticing $u_k^{(1)T}u_{k+1}^{(2)}=0$, we have

$$V_k^{(1)T}u_k^{(2)}=\big(u_{k-1}^{(1)}+a_{k-1}\big(Au_{k-1}^{(1)}-\big(u_{k-1}^{(1)T}Au_{k-1}^{(1)}\big)u_{k-1}^{(1)}\big)+O\big(a_{k-1}^2\big)+a_{k-1}\varepsilon_k^{(1)}\big)^Tu_k^{(2)}$$
$$=a_{k-1}\big(Au_{k-1}^{(1)}\big)^Tu_k^{(2)}+O\big(a_{k-1}^2\big)+a_{k-1}\varepsilon_k^{(1)T}u_k^{(2)}. \tag{22}$$

Since $d(u_k^{(1)},J_1)\xrightarrow[k\to\infty]{}0$ and $u_k^{(1)T}u_{k+1}^{(2)}=0$, continuing (22) for any $u^{(1)}\in J_1$ we have

$$V_k^{(1)T}u_k^{(2)}=a_{k-1}\big(Au_{k-1}^{(1)}\big)^Tu_k^{(2)}+O\big(a_{k-1}^2\big)+a_{k-1}\varepsilon_k^{(1)T}u_k^{(2)}$$
$$=a_{k-1}\big(Au_{k-1}^{(1)}-Au^{(1)}+Au^{(1)}-\lambda(1)u_{k-1}^{(1)}\big)^Tu_k^{(2)}+O\big(a_{k-1}^2\big)+a_{k-1}\varepsilon_k^{(1)T}u_k^{(2)}$$
$$=O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+O\big(a_k^2\big)+a_{k-1}\varepsilon_k^{(1)T}u_k^{(2)}, \tag{23}$$

which incorporating with (21) leads to

$$\big\|\tilde u_{k+1}^{(2)}\big\|^{-1}=1-a_ku_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}+O\big(a_k^2\big)$$

and a rough expression for $u_{k+1}^{(2)}$:

$$u_{k+1}^{(2)}=\big(P_k^{(1)}u_k^{(2)}+a_kP_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big)\cdot\big[1-a_ku_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}+O\big(a_k^2\big)\big]$$
$$=P_k^{(1)}u_k^{(2)}+a_kP_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}-a_k\big(u_k^{(2)T}P_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}\big)P_k^{(1)}u_k^{(2)}+O\big(a_k^2\big)$$
$$=P_k^{(1)}u_k^{(2)}+O(a_k). \tag{24}$$

Substituting this to the right-hand side of (23) gives

$$V_k^{(1)T}u_k^{(2)}=O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+O\big(a_k^2\big)+a_{k-1}\varepsilon_k^{(1)T}\big(P_{k-1}^{(1)}u_{k-1}^{(2)}+O(a_{k-1})\big)$$
$$=O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+O\big(a_k^2\big)+a_{k-1}\eta_k^{(1)}, \tag{25}$$

where $\eta_k^{(1)}\triangleq\varepsilon_k^{(1)T}P_{k-1}^{(1)}u_{k-1}^{(2)}$ is an $\mathcal{F}_k$-measurable mds.

This means that (19) is valid for $i=2$.

By noticing $u_k^{(1)}=u_{k-1}^{(1)}+O(a_{k-1})$ from (25) we obtain

$$V_k^{(1)}V_k^{(1)T}u_k^{(2)}=\big(u_{k-1}^{(1)}+O(a_{k-1})\big)\big(O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+O\big(a_k^2\big)+a_{k-1}\varepsilon_k^{(1)T}P_{k-1}^{(1)}u_{k-1}^{(2)}\big)$$
$$=O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+O\big(a_k^2\big)+a_{k-1}\gamma_k(2) \tag{26}$$

where $\gamma_k(2)\triangleq u_{k-1}^{(1)}\varepsilon_k^{(1)T}P_{k-1}^{(1)}u_{k-1}^{(2)}$ is an $\mathcal{F}_k$-measurable mds.

By taking (18) into account, this implies (20) for $i=2$.

By (26) we have

$$a_k\big(A_{k+1}V_k^{(1)}V_k^{(1)T}-V_k^{(1)}V_k^{(1)T}A_{k+1}V_k^{(1)}V_k^{(1)T}\big)u_k^{(2)}=O\big(a_{k-1}^2\big), \tag{27}$$

and for any $u^{(1)}(k)\in J_1$ by noticing $u_{k-1}^{(1)T}u_k^{(2)}=0$

$$a_kV_k^{(1)}V_k^{(1)T}A_{k+1}u_k^{(2)}=a_kV_k^{(1)}\big(Au_k^{(1)}\big)^Tu_k^{(2)}+a_kV_k^{(1)}V_k^{(1)T}N_{k+1}u_k^{(2)}$$
$$=a_kV_k^{(1)}\big(Au_k^{(1)}-Au^{(1)}(k)+\lambda(1)u^{(1)}(k)-\lambda(1)u_{k-1}^{(1)}+\lambda(1)u_{k-1}^{(1)}\big)^Tu_k^{(2)}$$
$$+a_kV_k^{(1)}V_k^{(1)T}N_{k+1}u_k^{(2)}$$
$$=O\big(a_kd\big(u_k^{(1)},J_1\big)\big)+a_kV_k^{(1)}V_k^{(1)T}N_{k+1}u_k^{(2)}. \tag{28}$$

We are now in a position to derive (8) for $i=2$.

Noticing $P_k^{(1)}=I-u_k^{(1)}u_k^{(1)T}$, for $k\geqslant 1$ we have

$$\tilde u_{k+1}^{(2)}=P_k^{(1)}u_k^{(2)}+a_kP_k^{(1)}A_{k+1}P_k^{(1)}u_k^{(2)}$$
$$=u_k^{(2)}-V_k^{(1)}V_k^{(1)T}u_k^{(2)}+a_k\big(I-V_k^{(1)}V_k^{(1)T}\big)A_{k+1}\big(I-V_k^{(1)}V_k^{(1)T}\big)u_k^{(2)}$$
$$=u_k^{(2)}+a_kA_{k+1}u_k^{(2)}-V_k^{(1)}V_k^{(1)T}u_k^{(2)}$$
$$-a_k\big(A_{k+1}V_k^{(1)}V_k^{(1)T}+V_k^{(1)}V_k^{(1)T}A_{k+1}-V_k^{(1)}V_k^{(1)T}A_{k+1}V_k^{(1)}V_k^{(1)T}\big)u_k^{(2)}. \tag{29}$$

By (26), (27), (28), and (29) for all sufficiently large k we have

$$\tilde{u}_{k+1}^{(2)} = u_k^{(2)} + a_k A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)}, \tag{30}$$

where $a_k \tilde{\delta}_{k+1}^{(2)} = a_k N_{k+1} u_k^{(2)} - a_{k-1}\gamma_k(2) - a_k V_k^{(1)} V_k^{(1)T} N_{k+1} u_k^{(2)}$ with $N_{k+1}u_k^{(2)}$, $\gamma_k(2)$, and $V_k^{(1)} V_k^{(1)T} N_{k+1} u_k^{(2)}$ being mds' measurable with respect to $\mathcal{F}_{k+1}$, $\mathcal{F}_k$, and $\mathcal{F}_{k+1}$, respectively.

Therefore, for large k we have

$$\begin{aligned}
\big\|\tilde{u}_{k+1}^{(2)}\big\|^{-1} &= \Big(\big(u_k^{(2)} + a_k A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)}\big)^T \\
&\quad \times \big(u_k^{(2)} + a_k A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)}\big)\Big)^{-\frac{1}{2}} \\
&= \big(1 + 2a_k u_k^{(2)T} A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)T} u_k^{(2)} + a_k u_k^{(2)T} \tilde{\delta}_{k+1}^{(2)}\big)^{-\frac{1}{2}} \\
&= 1 - a_k u_k^{(2)T} A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) - a_k \tilde{\delta}_{k+1}^{(2)T} u_k^{(2)},
\end{aligned} \tag{31}$$

and hence

$$\begin{aligned}
u_{k+1}^{(2)} &= \big(u_k^{(2)} + a_k A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)}\big) \\
&\quad \cdot \big(1 - a_k u_k^{(2)T} A u_k^{(2)} + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) - a_k \tilde{\delta}_{k+1}^{(2)T} u_k^{(2)}\big) \\
&= u_k^{(2)} + a_k\big(A u_k^{(2)} - \big(u_k^{(2)T} A u_k^{(2)}\big) u_k^{(2)}\big) + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \tilde{\delta}_{k+1}^{(2)} - a_k \tilde{\delta}_{k+1}^{(2)T} u_k^{(2)} u_k^{(2)}.
\end{aligned} \tag{32}$$

Here $\tilde{\delta}_{k+1}^{(2)T} u_k^{(2)} u_k^{(2)}$ is not an mds, but replacing $u_k^{(2)}$ with the expression given by (24) in $a_k \tilde{\delta}_{k+1}^{(2)T} u_k^{(2)} u_k^{(2)}$ of (32) leads to

$$u_{k+1}^{(2)} = u_k^{(2)} + a_k\big(A u_k^{(2)} - \big(u_k^{(2)T} A u_k^{(2)}\big) u_k^{(2)}\big) + O\big(a_k\big(a_k + d\big(u_k^{(1)}, J_1\big)\big)\big) + a_k \delta_{k+1}(2), \tag{33}$$

where $a_k \delta_{k+1}(2) = a_k \tilde{\delta}_{k+1}^{(2)} - a_k \tilde{\delta}_{k+1}^{(2)T} (P_{k-1}^{(1)} u_{k-1}^{(2)})(P_{k-1}^{(1)} u_{k-1}^{(2)})$. By the property of $a_k \tilde{\delta}_{k+1}^{(2)}$ mentioned above, $a_k \delta_{k+1}(2)$ is a linear combination of two mds' measurable with respect to $\mathcal{F}_{k+1}$ and $\mathcal{F}_k$. By A2 $\delta_{k+1}(2)$ is bounded, and hence $\sum_{k=1}^{\infty} a_k \delta_{k+1}(2) < \infty$.

Similar to (14)–(16), there exists $\lambda(2) \in V(J)$ such that

$$d\big(u_k^{(2)T} A u_k^{(2)}, \lambda(2)\big) \to 0 \quad \text{and} \quad d\big(u_k^{(2)}, J_2\big) \to 0, \tag{34}$$

where $J_2 \subset J$ is composed of limiting points of $\{u_k^{(2)}\}$ and

$$Au = \lambda(2) u, \quad \forall u \in J_2. \tag{35}$$

By (34) and (35) it follows that $d(A u_k^{(2)}, \lambda(2) u_k^{(2)}) \xrightarrow[k \to \infty]{} 0$.

Thus, we have shown that (18)–(20) are valid for $i = 2$, and the theorem is valid for $i = 1, 2$.

We now inductively prove (18)–(20) and the theorem. Assume that (18)–(20) and the theorem itself are valid for $i = 1, 2, \ldots, j \leqslant n-1$. We show that they are also true for $i = j+1$.

Noticing that the columns of $V_{k+1}^{(j)}$ are orthogonal, we have

$$\begin{aligned}
V_{k+1}^{(j)T} V_{k+1}^{(j)} &= \big[u_{k+1}^{(1)}, \ldots, P_{k+1}^{(j-1)} u_{k+1}^{(j)}\big]^T \big[u_{k+1}^{(1)}, \ldots, P_{k+1}^{(j-1)} u_{k+1}^{(j)}\big] \\
&= \begin{bmatrix} 1 & & & \\ & u_{k+1}^{(2)T} P_{k+1}^{(1)} u_{k+1}^{(2)} & & \\ & & \ddots & \\ & & & u_{k+1}^{(j)T} P_{k+1}^{(j-1)} u_{k+1}^{(j)} \end{bmatrix} \\
&= \begin{bmatrix} 1 & & & \\ & 1 - u_{k+1}^{(2)T} V_{k+1}^{(1)} V_{k+1}^{(1)+} u_{k+1}^{(2)} & & \\ & & \ddots & \\ & & & 1 - u_{k+1}^{(j)T} V_{k+1}^{(j-1)} V_{k+1}^{(j-1)+} u_{k+1}^{(j)} \end{bmatrix}.
\end{aligned}$$

From (18) and (19) for any $i = 1, \ldots, j$, we then have

$$\big(V_{k+1}^{(j)T} V_{k+1}^{(j)}\big)^{-1} = \big(I + O\big(a_k^2\big)\big)^{-1} = I + O\big(a_k^2\big). \tag{36}$$

This proves that (18) is true for $i = j+1$.

We now show that (19), (20) are also true for $i = j+1$.

Multiplying both sides of

$$\tilde{u}_{k+1}^{(j+1)} = P_k^{(j)} u_k^{(j+1)} + a_k P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}$$

by $V_k^{(j)T}$ from left we see $V_k^{(j)T} \tilde{u}_{k+1}^{(j+1)} = 0$, and hence

$$V_k^{(j)T} u_{k+1}^{(j+1)} = 0 \tag{37}$$

for sufficiently large k. From here it follows that

$$\begin{aligned} u_{k+1}^{(j+1)T} V_k^{(j)} &= u_{k+1}^{(j+1)T} \big[V_k^{(j-1)}, P_k^{(j-1)} u_k^{(j)} \big] \\ &= u_{k+1}^{(j+1)T} \big[V_k^{(i-1)}, P_k^{(i-1)} u_k^{(i)}, \ldots, P_k^{(j-1)} u_k^{(j)} \big] = 0. \end{aligned}$$

Thus, for any $i = 1, \ldots, j$ we have

$$0 = u_{k+1}^{(j+1)T} P_k^{(i-1)} u_k^{(i)} = u_{k+1}^{(j+1)T} u_k^{(i)} - u_{k+1}^{(j+1)T} V_k^{(i-1)} V_k^{(i-1)+} u_k^{(i)} = u_{k+1}^{(j+1)T} u_k^{(i)},$$

and hence

$$u_{k+1}^{(j+1)T} \big[u_k^{(1)}, \ldots, u_k^{(j)} \big] = 0, \quad k \geqslant 1. \tag{38}$$

By the inductive assumptions from (8) we have

$$\begin{aligned} u_{k+1}^{(i+1)} &= u_k^{(i+1)} + O\left(a_k \left(a_k + \sum_{s=1}^{i+1} d\big(u_k^{(s)}, J_s\big) \right) \right) + a_k \delta_{k+1}(i+1) \\ &= u_k^{(i+1)} + O(a_k), \quad i = 1, \ldots, j-1. \end{aligned} \tag{39}$$

By the inductive assumption,

$$V_{k+1}^{(i)} V_{k+1}^{(i)+} u_{k+1}^{(i+1)} = O\left(a_k \left(a_k + \sum_{s=1}^{i} d\big(u_k^{(s)}, J_s\big) \right) \right) + a_k \gamma_{k+1}(i+1), \quad i = 1, \ldots, j-1,$$

we have

$$\begin{aligned} V_{k+1}^{(j)} &= \big[u_{k+1}^{(1)}, P_{k+1}^{(1)} u_{k+1}^{(2)}, \ldots, P_{k+1}^{(j-1)} u_{k+1}^{(j)} \big] \\ &= \big[u_{k+1}^{(1)}, u_{k+1}^{(2)} - V_{k+1}^{(1)} V_{k+1}^{(1)+} u_{k+1}^{(2)}, \ldots, u_{k+1}^{(j)} - V_{k+1}^{(j-1)} V_{k+1}^{(j-1)+} u_{k+1}^{(j)} \big] \\ &= \big[u_{k+1}^{(1)}, u_{k+1}^{(2)}, \ldots, u_{k+1}^{(j)} \big] + O\left(a_k \left(a_k + \sum_{s=1}^{j-1} d\big(u_k^{(s)}, J_s\big) \right) \right) - a_k \big[0, \gamma_{k+1}(2), \ldots, \gamma_{k+1}(j) \big] \\ &= \big[u_{k+1}^{(1)}, u_{k+1}^{(2)}, \ldots, u_{k+1}^{(j)} \big] + O(a_k) \end{aligned} \tag{40}$$

and

$$\begin{aligned} u_{k+1}^{(j+1)T} P_{k+1}^{(i)} u_{k+1}^{(i+1)} &= u_{k+1}^{(j+1)T} u_{k+1}^{(i+1)} - u_{k+1}^{(j+1)T} V_{k+1}^{(i)} V_{k+1}^{(i)+} u_{k+1}^{(i+1)} \\ &= u_{k+1}^{(j+1)T} u_{k+1}^{(i+1)} + O\left(a_k \left(a_k + \sum_{s=1}^{i} d\big(u_k^{(s)}, J_s\big) \right) \right) + a_k u_{k+1}^{(j+1)T} \gamma_{k+1}(i+1) \\ &= u_{k+1}^{(j+1)T} \left(u_k^{(i+1)} + O\left(a_k \left(a_k + \sum_{s=1}^{i+1} d\big(u_k^{(s)}, J_s\big) \right) \right) + a_k \delta_{k+1}(i+1) \right) + a_k u_{k+1}^{(j+1)T} \gamma_{k+1}(i+1) \\ &= O\left(a_k \left(a_k + \sum_{s=1}^{i+1} d\big(u_k^{(s)}, J_s\big) \right) \right) + a_k u_{k+1}^{(j+1)T} \big(\delta_{k+1}(i+1) + \gamma_{k+1}(i+1) \big), \quad i = 1, \ldots, j-1, \end{aligned} \tag{41}$$

where for the last two equalities in (41), (38) and (39) are used.

Therefore, by (41) we obtain

$$u_{k+1}^{(j+1)T}V_{k+1}^{(j)} = u_{k+1}^{(j+1)T}\big[u_{k+1}^{(1)},\dots,P_{k+1}^{(j-1)}u_{k+1}^{(j)}\big]$$
$$= O\left(a_k\left(a_k + \sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k u_{k+1}^{(j+1)T}\big[0, \delta_{k+1}(2)+\gamma_{k+1}(2),\dots,\delta_{k+1}(j)+\gamma_{k+1}(j)\big]. \tag{42}$$

However, this is still not in the form of (19), because the last term in (42) is not expressed as linear combinations of mds'. Let us express $u_{k+1}^{(j+1)}$ via vectors of time k.

Noticing that $P_k^{(j)}P_k^{(j)} = P_k^{(j)}$ and $P_k^{(j)} = I - V_k^{(j)}V_k^{(j)+}$, for sufficiently large k we have

$$\begin{aligned}
\big\|\tilde{u}_{k+1}^{(j+1)}\big\|^{-1} &= \big\{\big[P_k^{(j)}u_k^{(j+1)} + a_k P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)}\big]^T\big[P_k^{(j)}u_k^{(j+1)} + a_k P_k^{(j)}\cdot A_{k+1}P_k^{(j)}u_k^{(j+1)}\big]\big\}^{-\frac{1}{2}}\\
&= \big[u_k^{(j+1)T}P_k^{(j)}u_k^{(j+1)} + 2a_k u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)} + a_k^2 u_k^{(j+1)T}\cdot P_k^{(j)}A_{k+1}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)}\big]^{-\frac{1}{2}}\\
&= \big[1 - u_k^{(j+1)T}V_k^{(j)}V_k^{(j)+}u_k^{(j+1)} + 2a_k u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)}\\
&\quad + a_k^2 u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)}\big]^{-\frac{1}{2}}\\
&= 1 + O\big(a_k^2\big) + \frac{1}{2}u_k^{(j+1)T}V_k^{(j)}\big(V_k^{(j)T}V_k^{(j)}\big)^{-1}V_k^{(j)T}u_k^{(j+1)} - a_k u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)},
\end{aligned} \tag{43}$$

which combining with (42) leads to

$$\big\|\tilde{u}_{k+1}^{(j+1)}\big\|^{-1} = 1 - a_k u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)} + O\big(a_k^2\big). \tag{44}$$

Therefore, for large enough k we have

$$\begin{aligned}
u_{k+1}^{(j+1)} &= \big(P_k^{(j)}u_k^{(j+1)} + a_k P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)}\big)\cdot\big[1 - a_k u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}\cdot u_k^{(j+1)} + O\big(a_k^2\big)\big]\\
&= P_k^{(j)}u_k^{(j+1)} + a_k P_k^{(j)}A_{k+1}P_k^{(j)}u_k^{(j+1)} - a_k\big(u_k^{(j+1)T}P_k^{(j)}A_{k+1}P_k^{(j)}\cdot u_k^{(j+1)}\big)P_k^{(j)}u_k^{(j+1)} + O\big(a_k^2\big)\\
&= P_k^{(j)}u_k^{(j+1)} + O(a_k),
\end{aligned} \tag{45}$$

and hence by (36) and (40)

$$\begin{aligned}
u_{k+1}^{(j+1)} &= P_k^{(j)}\cdot\big(P_{k-1}^{(j)}u_{k-1}^{(j+1)} + O(a_{k-1})\big) + O(a_k)\\
&= P_{k-1}^{(j)}u_{k-1}^{(j+1)} - V_k^{(j)}V_k^{(j)+}\big(P_{k-1}^{(j)}u_{k-1}^{(j+1)}\big) + O(a_k)\\
&= P_{k-1}^{(j)}u_{k-1}^{(j+1)} - \big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]\big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]^T P_{k-1}^{(j)}u_{k-1}^{(j+1)} + O(a_k).
\end{aligned} \tag{46}$$

Again by (36) and (40) we have

$$\begin{aligned}
P_{k-1}^{(j)}u_{k-1}^{(j+1)} &= P_{k-1}^{(j)}\cdot\big(P_{k-2}^{(j)}u_{k-2}^{(j+1)} + O(a_{k-2})\big)\\
&= P_{k-2}^{(j)}u_{k-2}^{(j+1)} - V_{k-1}^{(j)}V_{k-1}^{(j)+}P_{k-2}^{(j)}u_{k-2}^{(j+1)} + O(a_{k-2})\\
&= P_{k-2}^{(j)}u_{k-2}^{(j+1)} - \big(\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big] + O(a_{k-2})\big)\big(I + O\big(a_{k-2}^2\big)\big)\big(\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big] + O(a_{k-2})\big)^T\\
&\quad\cdot\big(P_{k-2}^{(j)}u_{k-2}^{(j+1)}\big) + O(a_{k-2})\\
&= P_{k-2}^{(j)}u_{k-2}^{(j+1)} - \big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]^T P_{k-2}^{(j)}u_{k-2}^{(j+1)} + O(a_{k-2})\\
&= \big(I - \big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]^T\big)P_{k-2}^{(j)}u_{k-2}^{(j+1)} + O(a_{k-2}).
\end{aligned} \tag{47}$$

Putting (47) into (46) yields

$$\begin{aligned}
u_{k+1}^{(j+1)} &= \big(I - \big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]\big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]^T\big)\big(I - \big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]^T\big)P_{k-2}^{(j)}u_{k-2}^{(j+1)} + O(a_k)\\
&= \big(I - \big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]\big[u_{k-1}^{(1)},\dots,u_{k-1}^{(j)}\big]^T\big)\big(I - \big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]\big[u_{k-2}^{(1)},\dots,u_{k-2}^{(j)}\big]^T\big)\\
&\quad\cdots\big(I - \big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}\big]\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}\big]^T\big)P_{k-j}^{(j)}u_{k-j}^{(j+1)} + O(a_k).
\end{aligned} \tag{48}$$

By (39) $u_{k+1}^{(i+1)} = u_k^{(i+1)} + O(a_k) = u_{k-j}^{(i+1)} + O(a_k)$, $i = 1,\dots,j-1$, (48) can be rewritten as

$$u_{k+1}^{(j+1)} = \big(I - \big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}\big]\big[u_{k-1}^{(1)},\dots,u_{k-j}^{(j)}\big]^T\big)^j P_{k-j}^{(j)}u_{k-j}^{(j+1)} + O(a_k). \tag{49}$$

Putting this into the right-hand side of (41) yields

$$\begin{aligned} u_{k+1}^{(j+1)T} P_{k+1}^{(i)} u_{k+1}^{(i+1)} &= O\left(a_k\left(a_k+\sum_{s=1}^{i+1} d\big(u_k^{(s)}, J_s\big)\right)\right) \\ &\quad + a_k\big(\big(I-\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}\big]\big[u_{k-1}^{(1)},\dots,u_{k-j}^{(j)}\big]^T\big)^j P_{k-j}^{(j)} u_{k-j}^{(j+1)} + O(a_k)\big)^T \\ &\quad \times \big(\delta_{k+1}(i+1)+\gamma_{k+1}(i+1)\big) \\ &= O\left(a_k\left(a_k+\sum_{s=1}^{i+1} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k \eta_{k+1}^{(j)}(i+1), \quad i=1,\dots,j-1, \end{aligned} \tag{50}$$

where $\eta_{k+1}^{(j)}(i+1) \triangleq u_{k-j}^{(j+1)T} P_{k-j}^{(j)}(I-[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}][u_{k-1}^{(1)},\dots,u_{k-j}^{(j)}]^T)^j(\delta_{k+1}(i+1)+\gamma_{k+1}(i+1))$.

By the inductive assumptions, from (8) it is seen that $\delta_{k+1}(j)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,$ and $\mathcal{F}_{k+2-j}$, respectively, and $\gamma_{k+1}(j)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,$ and $\mathcal{F}_{k+3-j}$, respectively. Therefore, $\eta_{k+1}^{(j)}(i+1)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,$ and $\mathcal{F}_{k+2-j}$, respectively.

From (50) it follows that

$$\begin{aligned} u_{k+1}^{(j+1)T} V_{k+1}^{(j)} &= u_{k+1}^{(j+1)T}\big[u_{k+1}^{(1)},\dots,P_{k+1}^{(j-1)} u_{k+1}^{(j)}\big] \\ &= O\left(a_k\left(a_k+\sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k\eta_{k+1}^{(j)}, \end{aligned} \tag{51}$$

where $\eta_{k+1}^{(j)} \triangleq [\eta_{k+1}^{(1)}(1),\eta_{k+1}^{(j)}(2),\dots,\eta_{k+1}^{(j)}(j)]$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,$ and $\mathcal{F}_{k+2-j}$, respectively, where $\eta_{k+1}^{(1)}(1) \triangleq \eta_{k+1}^{(1)}$. Thus, (19) is proved for $i=j+1$.

By (36) and (40) from (51) it follows that

$$\begin{aligned} V_{k+1}^{(j)} V_{k+1}^{(j)+} u_{k+1}^{(j+1)} &= V_{k+1}^{(j)}\big(V_{k+1}^{(j)T} V_{k+1}^{(j)}\big)^{-1} V_{k+1}^{(j)T} u_{k+1}^{(j+1)} \\ &= \big(\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}\big]+O(a_k)\big)\big(I+O\big(a_k^2\big)\big)\left(O\left(a_k\left(a_k+\sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right)+a_k\eta_{k+1}^{(j)}\right)^T \\ &= O\left(a_k\left(a_k+\sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k\gamma_{k+1}(j+1), \end{aligned} \tag{52}$$

where $\gamma_{k+1}(j+1) \triangleq [u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}]\eta_{k+1}^{(j)T}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,$ and $\mathcal{F}_{k+2-j}$, respectively.

This means that (20) is valid for $i=j+1$.

We are now in a position to show that the theorem is true for $i=j+1$ either.

By (40) we then have

$$\begin{aligned} & a_k V_k^{(j)} V_k^{(j)+} A_{k+1} u_k^{(j+1)} \\ &\quad = a_k V_k^{(j)}\big(V_k^{(j)T} V_k^{(j)}\big)^{-1} V_k^{(j)T} A u_k^{(j+1)} + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)} \\ &\quad = a_k V_k^{(j)}\big(V_k^{(j)T} V_k^{(j)}\big)^{-1}\big(A\big[u_k^{(1)},\dots,u_k^{(j)}\big]\big)^T u_k^{(j+1)} + O\big(a_k^2\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)} \end{aligned} \tag{53}$$

for sufficiently large k.

By inductive assumptions, for any $i \in \{1,\dots,j\}$ there is a sequence $\{u^{(i)}(k)\} \in J_i$ such that $d(u_k^{(i)}, u^{(i)}(k)) \xrightarrow[k\to\infty]{} 0$ and $Au^{(i)}(k) = \lambda(i)u^{(i)}(k)$.

By (38), (39) we have

$$\begin{aligned} & a_k V_k^{(j)}\big(V_k^{(j)T} V_k^{(j)}\big)^{-1}\big(Au_k^{(i)}\big)^T u_k^{(j+1)} + O\big(a_k^2\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)} \\ &\quad = a_k V_k^{(j)}\big(V_k^{(j)T} V_k^{(j)}\big)^{-1}\cdot\big(Au_k^{(i)} - Au^{(i)}(k) + \lambda(i)u^{(i)}(k) - \lambda(i)u_k^{(i)}(k) + \lambda(i)u_k^{(i)}\big)^T u_k^{(j+1)} \\ &\qquad + O\big(a_k^2\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)} \\ &\quad = \lambda(i) a_k V_k^{(j)}\big(V_k^{(j)T} V_k^{(j)}\big)^{-1} u_k^{(i)T} u_k^{(j+1)} + O\big(a_k\big(a_k + d\big(u_k^{(i)}, J_i\big)\big)\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)} \end{aligned}$$

$$= \lambda(i) a_k V_k^{(j)} \big(V_k^{(j)T} V_k^{(j)}\big)^{-1} \big(u_{k-1}^{(i)} + O(a_k)\big)^T u_k^{(j+1)} + O\big(a_k\big(a_k + d\big(u_k^{(i)}, J_i\big)\big)\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)}$$
$$= O\big(a_k\big(a_k + d\big(u_k^{(i)}, J_i\big)\big)\big) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)}, \quad \forall i = 1, \ldots, j. \tag{54}$$

Putting the expression given by (54) into (53) yields

$$a_k V_k^{(j)} V_k^{(j)+} A_{k+1} u_k^{(j+1)} = O\left(a_k\left(a_k + \sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)}. \tag{55}$$

Then we have

$$\begin{aligned}
\tilde{u}_{k+1}^{(j+1)} &= P_k^{(j)} u_k^{(j+1)} + a_k P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} \\
&= \big(I - V_k^{(j)} V_k^{(j)+}\big) u_k^{(j+1)} + a_k \big(I - V_k^{(j)} V_k^{(j)+}\big) A_{k+1} \big(I - V_k^{(j)} V_k^{(j)+}\big) u_k^{(j+1)} \\
&= u_k^{(j+1)} + a_k A_{k+1} u_k^{(j+1)} - V_k^{(j)} V_k^{(j)+} u_k^{(j+1)} \\
&\quad - a_k \big(A_{k+1} V_k^{(j)} V_k^{(j)+} + V_k^{(j)} V_k^{(j)+} A_{k+1} - V_k^{(j)} V_k^{(j)+} A_{k+1} V_k^{(j)} V_k^{(j)+}\big) u_k^{(j+1)}.
\end{aligned} \tag{56}$$

Notice that (52) implies that

$$a_k \big(A_{k+1} V_k^{(j)} V_k^{(j)+} - V_k^{(j)} V_k^{(j)+} A_{k+1} V_k^{(j)} V_k^{(j)+}\big) u_k^{(j+1)} = O\big(a_k^2\big).$$

From (56) by (52) and (55) it follows that

$$\tilde{u}_{k+1}^{(j+1)} = u_k^{(j+1)} + a_k A u_k^{(j+1)} + O\left(a_k\left(a_k + \sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k \tilde{\delta}_{k+1}^{(j+1)}, \tag{57}$$

where $a_k \tilde{\delta}_{k+1}^{(j+1)} = a_k N_{k+1} u_k^{(j+1)} - a_{k-1} \gamma_k(j+1) - a_k V_k^{(j)} V_k^{(j)+} N_{k+1} u_k^{(j+1)}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1}, \mathcal{F}_k, \ldots,$ and $\mathcal{F}_{k+1-j}$, respectively.

An analysis similar to (31), (32), and (33) leads to (8) for $i = j+1$:

$$u_{k+1}^{(j+1)} = u_k^{(j+1)} + a_k \big(A u_k^{(j+1)} - \big(u_k^{(j+1)T} A u_k^{(j+1)}\big) u_k^{(j+1)}\big) + O\left(a_k\left(a_k + \sum_{s=1}^{j} d\big(u_k^{(s)}, J_s\big)\right)\right) + a_k \delta_{k+1}(j+1), \tag{58}$$

where $\delta_{k+1}(j+1)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1}, \mathcal{F}_k, \ldots,$ and $\mathcal{F}_{k+1-j}$, respectively.

Similar to (14), (15), (16) it is shown that there exists a $J_{j+1} \subset J$ such that

$$\lim_{k \to \infty} d\big(u_k^{(j+1)}, J_{j+1}\big) = 0, \tag{59}$$

where J_{j+1} is composed of limiting points of $\{u_k^{(j+1)}\}$.

Correspondingly, there is a $\lambda(j+1) \in V(J)$ such that

$$d\big(u_k^{(j+1)T} A u_k^{(j+1)}, \lambda(j+1)\big) \to 0, \quad \text{and} \quad Au = \lambda(j+1) u, \quad \forall u \in J_{j+1}, \tag{60}$$

which incorporating with (59) implies $d(A u_k^{(j+1)}, \lambda(j+1) u_k^{(j+1)}) \xrightarrow[k \to \infty]{} 0$.

Thus, (i), (ii), and (iii) of the theorem have been proved by induction.

Finally, (6) can be rewritten as

$$\lambda_{k+1}^{(j)} = \lambda_k^{(j)} - a_k \big(\lambda_k^{(j)} - \lambda(j) + \varepsilon_{k+1}^{(j)}\big),$$

where

$$\varepsilon_{k+1}^{(j)} \triangleq \lambda(j) - u_k^{(j)T} A u_k^{(j)} - u_k^{(j)T} N_{k+1} u_k^{(j)}.$$

Since $\lambda(j) - u_k^{(j)T} A u_k^{(j)} \to 0$ and $\sum_{k=1}^{\infty} a_k u_k^{(j)T} N_{k+1} u_k^{(j)} < \infty$, the conclusion (iv) follows from Lemma 3.1.1 of [3]. □

4. Ordered convergence of estimates

In this section we show that convergence established in Theorem 1 is actually ordered in the sense that $\lambda(i) = \lambda^{(i)}$, where the eigenvalues $\{\lambda^{(i)}\}$ are ordered: $\lambda^{(1)} \geqslant \lambda^{(2)} \geqslant \cdots \geqslant \lambda^{(n)}$. But, we first give two lemmas.

Lemma 1. *Assume the random sequence $\{X_k,\ k \geqslant 0\}$ is generated by the following recursion*

$$X_{k+1} = X_k + a_k\alpha_k X_k + a_k\varepsilon_{k+1} + O\big(a_k^2\big), \tag{61}$$

where the real sequence $\{a_k\}$ is such that $a_k > 0$, $\sum_k a_k = \infty$, $\sum_k a_k^2 < \infty$; the real number α_k has a positive limit: $\alpha_k \xrightarrow[k\to\infty]{} \alpha > 0$; and $\varepsilon_k = \sum_{i=1}^{l} \varepsilon_{k-i+1}^{(i)}$, $l \in [1,\infty)$, where $\{\varepsilon_k^{(i)}, \mathcal{F}_k\}$ is an mds for any $i \in [1,l]$. Moreover, $\liminf_k E\{\|\varepsilon_{k+1}^{(1)}\| \mid \mathcal{F}_k\} > 0$, and $\varepsilon_k^{(i)} \xrightarrow[k\to\infty]{} 0$ for any $i \geqslant 2$.

Then $P(X_k \to 0) = 0$.

The proof is given in Appendix A.

Lemma 2. *Assume* A1 *and* A2 *hold. Then*

$$V_k^{(i)} V_k^{(i)+} A P_k^{(i)} = V_k^{(i)} V_k^{(i)+} A\big(I - V_k^{(i)} V_k^{(i)+}\big) = o(1), \tag{62}$$

$$V_{k+1}^{(i)} V_{k+1}^{(i)+} P_k^{(i)} = V_{k+1}^{(i)} V_{k+1}^{(i)+} - V_{k+1}^{(i)} V_{k+1}^{(i)+} V_k^{(i)} V_k^{(i)+} = o(a_k) + a_k \varepsilon_{k+1}^{(i+1)'}, \tag{63}$$

where $V_k^{(i)}$, $i = 1,\dots,n$ are given by (1)–(3)*, and $\varepsilon_{k+1}^{(i+1)'}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1}, \mathcal{F}_k, \dots, \mathcal{F}_{k+2-i}$, respectively and each mds is with one of $u_{k-i}^{(1)}, \dots, u_{k-i}^{(i)}$ as its left factor.*

Proof. By (18), (62) and (63) are equivalent to the following expressions

$$V_k^{(i)} V_k^{(i)T} A\big(I - V_k^{(i)} V_k^{(i)T}\big) = o(1), \tag{64}$$

$$V_{k+1}^{(i)} V_{k+1}^{(i)T} - V_{k+1}^{(i)} V_{k+1}^{(i)T} V_k^{(i)} V_k^{(i)T} = o(a_k) + a_k \varepsilon_{k+1}^{(i+1)'}. \tag{65}$$

Let us prove (64) and (65) by induction.

For $i = 1$, by Theorem 1 $d(u_k^{(1)}, J_1) \xrightarrow[k\to\infty]{} 0$ a.s. and $u_k^{(1)T} A - \lambda^{(j)} u^{(1)T} = (u_k^{(1)T} - u^{(1)T})A$ for some j and any $u^{(1)} \in J_1$. Assume that $u^{(1)}(k) \in J_1$ and $d(u_k^{(1)}, J_1) = \|u_k^{(1)} - u^{(1)}(k)\|$. Consequently, by noticing $u^{(1)T}(k) u^{(1)}(k) = 1$, we have

$$\begin{aligned}
&u_k^{(1)} u_k^{(1)T} A - u_k^{(1)} u_k^{(1)T} A u_k^{(1)} u_k^{(1)T} \\
&\quad = u_k^{(1)}\big(u_k^{(1)T} A - \lambda^{(j)} u^{(1)T}(k)\big) + \lambda^{(j)} u_k^{(1)} u^{(1)T}(k) - u_k^{(1)}\big(u_k^{(1)T} A - \lambda^{(j)} u^{(1)T}(k)\big) u_k^{(1)} u_k^{(1)T} \\
&\qquad - \lambda^{(j)} u_k^{(1)} u^{(1)T}(k) u_k^{(1)} u_k^{(1)T} = O\big(\big\|u_k^{(1)} - u^{(1)}(k)\big\|\big) = O\big(d\big(u_k^{(1)}, J_1\big)\big) = o(1).
\end{aligned} \tag{66}$$

Thus, (64) holds for $i = 1$. We now show that (65) also takes place for $i = 1$.

By (9) it follows that

$$\begin{aligned}
u_{k+1}^{(1)} u_{k+1}^{(1)T} &= \big[u_k^{(1)} + a_k\big(A_{k+1} u_k^{(1)} - \big(u_k^{(1)T} A_{k+1} u_k^{(1)}\big) u_k^{(1)}\big) + O\big(a_k^2\big)\big] \\
&\quad \cdot \big[u_k^{(1)} + a_k\big(A_{k+1} u_k^{(1)} - \big(u_k^{(1)T} A_{k+1} u_k^{(1)}\big) u_k^{(1)}\big) + O\big(a_k^2\big)\big]^T \\
&= u_k^{(1)} u_k^{(1)T} + a_k A_{k+1} u_k^{(1)} u_k^{(1)T} - 2a_k\big(u_k^{(1)T} A_{k+1} u_k^{(1)}\big) u_k^{(1)} u_k^{(1)T} + a_k u_k^{(1)} u_k^{(1)T} A_{k+1} + O\big(a_k^2\big).
\end{aligned}$$

Consequently,

$$\begin{aligned}
&u_{k+1}^{(1)} u_{k+1}^{(1)T} - u_{k+1}^{(1)} u_{k+1}^{(1)T} u_k^{(1)} u_k^{(1)T} \\
&\quad = \big[u_k^{(1)} u_k^{(1)T} + a_k A_{k+1} u_k^{(1)} u_k^{(1)T} - 2a_k\big(u_k^{(1)T} A_{k+1} u_k^{(1)}\big) u_k^{(1)} u_k^{(1)T} + a_k u_k^{(1)} u_k^{(1)T} A_{k+1} + O\big(a_k^2\big)\big]\big(I - u_k^{(1)} u_k^{(1)T}\big) \\
&\quad = a_k u_k^{(1)} u_k^{(1)T} A_{k+1} - a_k u_k^{(1)} u_k^{(1)T} A_{k+1} u_k^{(1)} u_k^{(1)T} + O\big(a_k^2\big) \\
&\quad = a_k\big(u_k^{(1)} u_k^{(1)T} A - u_k^{(1)} u_k^{(1)T} A u_k^{(1)} u_k^{(1)T}\big) + O\big(a_k^2\big) + a_k \varepsilon_{k+1}^{(2)'},
\end{aligned} \tag{67}$$

where $\varepsilon_{k+1}^{(2)'} = u_k^{(1)} u_k^{(1)T} N_{k+1} - u_k^{(1)} u_k^{(1)T} N_{k+1} u_k^{(1)} u_k^{(1)T}$ is an $\mathcal{F}_{k+1}$-measurable mds and is with $u_k^{(1)}$ as its left factor.

By (66) from (67) it follows that

$$V_{k+1}^{(1)}V_{k+1}^{(1)T}-V_{k+1}^{(1)}V_{k+1}^{(1)T}V_k^{(1)}V_k^{(1)T}=o(a_k)+a_k\varepsilon_{k+1}^{(2)'}. \tag{68}$$

Thus, (65) holds for $i=1$.

Assume that for all sufficiently large k (64), (65) hold for $i=1,2,\dots,j-1$, $j\geqslant 2$.

We now show that (64) and (65) are also true for $i=j\leqslant n$. Let us first verify (64) for $i=j$.

From (20) it is seen that $V_k^{(j)}=[V_k^{(j-1)},P_k^{(j-1)}u_k^{(j)}]=[V_k^{(j-1)},u_k^{(j)}+o(a_{k-1})+a_{k-1}\gamma_k(j)]$, and hence

$$\begin{aligned}V_k^{(j)}V_k^{(j)T}&=\big[V_k^{(j-1)},u_k^{(j)}+o(a_{k-1})+a_{k-1}\gamma_k(j)\big]\big[V_k^{(j-1)},u_k^{(j)}+o(a_{k-1})+a_{k-1}\gamma_k(j)\big]^T\\&=V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}+o(a_{k-1})+a_{k-1}\gamma_k(j)u_k^{(j)T}+a_{k-1}u_k^{(j)}\gamma_k^T(j)\\&=V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}+O(a_k).\end{aligned} \tag{69}$$

Consequently,

$$\begin{aligned}&V_k^{(j)}V_k^{(j)T}A\big(I-V_k^{(j)}V_k^{(j)T}\big)\\&\quad=V_k^{(j)}V_k^{(j)T}A-V_k^{(j)}V_k^{(j)T}AV_k^{(j)}V_k^{(j)T}\\&\quad=\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)A-\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)A\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)+O(a_k)\\&\quad=V_k^{(j-1)}V_k^{(j-1)T}A-V_k^{(j-1)}V_k^{(j-1)T}AV_k^{(j-1)}V_k^{(j-1)T}-V_k^{(j-1)}V_k^{(j-1)T}Au_k^{(j)}u_k^{(j)T}\\&\qquad-u_k^{(j)}u_k^{(j)T}AV_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}A-u_k^{(j)}u_k^{(j)T}Au_k^{(j)}u_k^{(j)T}+O(a_k).\end{aligned} \tag{70}$$

We want to show that the right-hand side of the above expression is $o(1)$. First, by the inductive assumption, the first two terms at the right-hand side of (70) give $o(1)$. Second, for its third term by Theorem 1, $d(Au_k^{(j)},\lambda^{(m)}u_k^{(j)})\xrightarrow[k\to\infty]{}0$ for some m and by (19) it follows that

$$V_k^{(j-1)}V_k^{(j-1)T}Au_k^{(j)}u_k^{(j)T}=\lambda^{(m)}V_k^{(j-1)}V_k^{(j-1)T}u_k^{(j)}u_k^{(j)T}+o(1)=o(1). \tag{71}$$

Similar to (71) we can show

$$u_k^{(j)}u_k^{(j)T}AV_k^{(j-1)}V_k^{(j-1)T}=\lambda^{(m)}u_k^{(j)}u_k^{(j)T}V_k^{(j-1)}V_k^{(j-1)T}+o(1)=o(1). \tag{72}$$

Finally, by Theorem 1 $u_k^{(j)T}A-(u_k^{(j)T}Au_k^{(j)})u_k^{(j)T}\to 0$, hence

$$u_k^{(j)}u_k^{(j)T}A-u_k^{(j)}u_k^{(j)T}Au_k^{(j)}u_k^{(j)T}=u_k^{(j)}\big(u_k^{(j)T}A-\big(u_k^{(j)T}Au_k^{(j)}\big)u_k^{(j)T}\big)=o(1). \tag{73}$$

Thus, we have shown that (64) is true for $i=j$.

We now prove that (65) is true for $i=j$.

Using the expression of $V_k^{(j)}V_k^{(j)T}$ given after the second equality in (69) we derive

$$\begin{aligned}&V_{k+1}^{(j)}V_{k+1}^{(j)T}-V_{k+1}^{(j)}V_{k+1}^{(j)T}V_k^{(j)}V_k^{(j)T}\\&\quad=V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}-\big(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}\big)\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)+o(a_k)\\&\qquad+a_k\gamma_{k+1}(j)u_{k+1}^{(j)T}+a_ku_{k+1}^{(j)}\gamma_{k+1}^T(j)-a_k\gamma_{k+1}(j)u_{k+1}^{(j)T}\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)-a_ku_{k+1}^{(j)}\gamma_{k+1}^T(j)\\&\qquad\cdot\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)-a_{k-1}\big(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}\big)\gamma_k(j)u_k^{(j)T}\\&\qquad-a_{k-1}\big(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}\big)u_k^{(j)}\gamma_k(j)^T\\&\quad=V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}-V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}V_k^{(j-1)}V_k^{(j-1)T}-u_{k+1}^{(j)}u_{k+1}^{(j)T}V_k^{(j-1)}V_k^{(j-1)T}-V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}u_k^{(j)}u_k^{(j)T}\\&\qquad+u_{k+1}^{(j)}u_{k+1}^{(j)T}-u_{k+1}^{(j)}u_{k+1}^{(j)T}u_k^{(j)}u_k^{(j)T}+o(a_k)+a_k\gamma_{k+1}(j)u_{k+1}^{(j)T}+a_ku_{k+1}^{(j)}\gamma_{k+1}(j)^T\\&\qquad-a_k\gamma_{k+1}(j)u_{k+1}^{(j)T}\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)-a_ku_{k+1}^{(j)}\gamma_{k+1}(j)^T\big(V_k^{(j-1)}V_k^{(j-1)T}+u_k^{(j)}u_k^{(j)T}\big)\\&\qquad-a_{k-1}\big(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}\big)\gamma_ku_k^{(j)T}-a_{k-1}\big(V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}\big)u_k^{(j)}\gamma_k(j)^T.\end{aligned} \tag{74}$$

Noting $u_{k+1}^{(i)}=u_k^{(i)}+o(1)=u_{k-j}^{(i)}+o(1)$ by (39) and $V_{k+1}^{(j)}=[u_{k-j}^{(1)},\dots,u_{k-j}^{(j)}]+o(1)$ by (40), we can rewrite (74) as

$$\begin{aligned}
&V_{k+1}^{(j)}V_{k+1}^{(j)T}-V_{k+1}^{(j)}V_{k+1}^{(j)T}V_k^{(j)}V_k^{(j)T}\\
&=V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}-V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}V_k^{(j-1)}V_k^{(j-1)T}-u_{k+1}^{(j)}u_{k+1}^{(j)T}V_k^{(j-1)}V_k^{(j-1)T}\\
&\quad-V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}u_k^{(j)}u_k^{(j)T}+u_{k+1}^{(j)}u_{k+1}^{(j)T}-u_{k+1}^{(j)}u_{k+1}^{(j)T}u_k^{(j)}u_k^{(j)T}+o(a_k)+a_k\gamma_{k+1}(j)u_{k-j}^{(j)T}\\
&\quad+a_ku_{k-j}^{(j)}\gamma_{k+1}(j)^T-a_k\gamma_{k+1}(j)u_{k-j}^{(j)T}\big(\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]^T+u_{k-j}^{(j)}u_{k-j}^{(j)T}\big)\\
&\quad-a_ku_{k-j}^{(j)}\gamma_{k+1}(j)^T\big(\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]^T+u_{k-j}^{(j)}u_{k-j}^{(j)T}\big)-a_{k-1}\\
&\quad\cdot\big(\big[u_{k-j-1}^{(1)},\dots,u_{k-j-1}^{(j-1)}\big]\big[u_{k-j-1}^{(1)},\dots,u_{k-j-1}^{(j-1)}\big]^T+u_{k-j-1}^{(j)}u_{k-j-1}^{(j)T}\big)\gamma_k(j)u_{k-j-1}^{(j)T}-a_{k-1}\\
&\quad\cdot\big(\big[u_{k-j-1}^{(1)},\dots,u_{k-j-1}^{(j-1)}\big]\big[u_{k-j-1}^{(1)},\dots,u_{k-j-1}^{(j-1)}\big]^T+u_{k-j-1}^{(j)}u_{k-j-1}^{(j)T}\big)u_{k-j-1}^{(j)}\gamma_k(j)^T. \qquad (75)
\end{aligned}$$

We want to express the right-hand side of (75) in the form of the right-hand side of (65) for $i=j$. Let us analyze each term at the right-hand side of (75).

First, the first two terms are estimated by the inductive assumption:

$$V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}-V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}V_k^{(j-1)}V_k^{(j-1)T}=o(a_k)+a_k\varepsilon_{k+1}^{(j)'}, \qquad (76)$$

where $\varepsilon_{k+1}^{(j)'}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+3-j}$, respectively, and each mds is with one of the vectors $u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}$ as its left factor.

Second, by (37) its third term equals zero.

Replacing $u_{k+1}^{(j)}$ in (19) with the expression given by the first equality of (39), we derive

$$\begin{aligned}
V_{k+1}^{(j-1)T}u_k^{(j)}&=o(a_k)+a_k\eta_{k+1}^{(j-1)}-a_kV_{k+1}^{(j-1)T}\delta_{k+1}(j)\\
&=o(a_k)+a_k\eta_{k+1}^{(j-1)}-a_k\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]^T\delta_{k+1}(j), \qquad (77)
\end{aligned}$$

where for the last equality (40) is used.

By (39), (40), and (77) it follows that

$$\begin{aligned}
&V_{k+1}^{(j-1)}V_{k+1}^{(j-1)T}u_k^{(j)}u_k^{(j)T}\\
&=\big(\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]+o(1)\big)\big(o(a_k)+a_k\eta_{k+1}^{(j-1)}-a_k\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]^T\delta_{k+1}(j)\big)\cdot\big(u_{k-j}^{(j)}+o(1)\big)^T\\
&=o(a_k)+a_k\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]\eta_{k+1}^{(j-1)}u_{k-j}^{(j)T}-a_k\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]\cdot\big[u_{k-j}^{(1)},\dots,u_{k-j}^{(j-1)}\big]^T\delta_{k+1}(j)u_{k-j}^{(j)T}. \qquad (78)
\end{aligned}$$

This gives the required expression for the fourth term at the right-hand side of (75).

Finally, by the first equality in (39) it follows that

$$\begin{aligned}
u_{k+1}^{(j)}u_{k+1}^{(j)T}&=\big(u_k^{(j)}+o(a_k)+a_k\delta_{k+1}(j)\big)\cdot\big(u_k^{(j)}+o(a_k)+a_k\delta_{k+1}(j)\big)^T\\
&=u_k^{(j)}u_k^{(j)T}+o(a_k)+a_k\delta_{k+1}(j)u_k^{(j)T}+a_ku_k^{(j)}\delta_{k+1}(j)^T, \qquad (79)
\end{aligned}$$

and hence

$$\begin{aligned}
&u_{k+1}^{(j)}u_{k+1}^{(j)T}-u_{k+1}^{(j)}u_{k+1}^{(j)T}u_k^{(j)}u_k^{(j)T}\\
&=\big[u_k^{(j)}u_k^{(j)T}+o(a_k)+a_k\delta_{k+1}(j)u_k^{(j)T}+a_ku_k^{(j)}\delta_{k+1}(j)^T\big]\big(I-u_k^{(j)}u_k^{(j)T}\big)\\
&=o(a_k)+a_ku_k^{(j)}\delta_{k+1}(j)^T-a_ku_k^{(j)}\delta_{k+1}(j)^Tu_k^{(j)}u_k^{(j)T}\\
&=o(a_k)+a_ku_{k-j}^{(j)}\delta_{k+1}(j)^T-a_ku_{k-j}^{(j)}\delta_{k+1}(j)^Tu_{k-j}^{(j)}u_{k-j}^{(j)T}, \qquad (80)
\end{aligned}$$

where the last equality is because $u_{k+1}^{(j)}=u_k^{(j)}+o(1)=u_{k-j}^{(j)}+o(1)$ by (39).

This gives expression for the fifth and sixth terms at the right-hand side of (75).

Notice that in (75), (78), and (80) $\eta_{k+1}^{(j-1)}$, $\delta_{k+1}(j)$, and $\gamma_{k+1}(j)$ are involved. They are the linear combinations of mds'. To be precise, $\eta_{k+1}(j)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+3-j}$, $\delta_{k+1}(j)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+2-j}$, and $\gamma_{k+1}(j)$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+3-j}$.

Thus, putting (76), (78), and (80) into (75) leads to

$$V_{k+1}^{(j)}V_{k+1}^{(j)T}-V_{k+1}^{(j)}V_{k+1}^{(j)T}V_k^{(j)}V_k^{(j)T}=o(a_k)+a_k\varepsilon_{k+1}^{(j+1)'},$$

where $\varepsilon_{k+1}^{(j+1)'}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1},\mathcal{F}_k,\dots,\mathcal{F}_{k+2-j}$.

It is noticed that for each terms containing either $\eta_{k+1}^{(j-1)}$ or $\delta_{k+1}(j)$ there is a left factor from $u_{k-j}^{(1)}, \dots, u_{k-j}^{(j)}$, while by definition, the expression of $\gamma_{k+1}(j)$ given immediately after (52) also includes a left factor $[u_{k-j}^{(1)}, \dots, u_{k-j}^{(j)}]$.

Thus, we have shown that (65) is true for $i = j$ and have completed the proof of the lemma. $\square$

To establish the ordered convergence of $\{u_k^{(i)}\}$ we need one more assumption A3 in addition to A1 and A2.

A3.

$$\liminf_n E\{\|x_k^T N_{k+1} y_k\| \mid \mathcal{F}_k\} > 0$$

for any $\mathcal{F}_k$-measurable x_k and y_k bounded from above and from zero:

$$0 < \liminf_{k\to\infty} \|x_k\| \leqslant \limsup_{k\to\infty} \|x_k\| < \infty, \qquad 0 < \liminf_{k\to\infty} \|y_k\| \leqslant \limsup_{k\to\infty} \|y_k\| < \infty. \tag{81}$$

Remark 2. Condition A3 has excluded the case $A_{k+1} \equiv A$ from consideration. As a matter of fact, in order to achieve the desired limit some observation noise is necessary, otherwise, it may happen that the algorithm is stuck at an undesired vector. To see this, let $A_{k+1} \equiv A$ and let the initial value $u_1^{(1)} \triangleq u^{(i)}$, where $u^{(i)}$ is a unit eigenvector of A corresponding to an eigenvalue $\lambda^{(i)}$ different from the largest one. Then,

$$\tilde{u}_{k+1}^{(1)} = u^{(i)} + a_k A u^{(i)} = (1 + a_k \lambda^{(i)}) u^{(i)},$$
$$u_{k+1}^{(1)} = u^{(i)},$$

and the algorithm will never converge to the desired $u^{(1)}$.

The following proposition gives sufficient conditions on $\{N_k\}$ in order A3 to be satisfied.

Proposition. *Assume that $(N_k, \mathcal{F}_k)$ with $N_k = \{N_{ij}(k)\}$ is a martingale difference sequence, $E(N_{ij}(k+1)N_{st}(k+1) \mid \mathcal{F}_k) = 0$ whenever $(ij) \neq (st)$, $\liminf_{k\to\infty} E(N_{ij}^2(k+1) \mid \mathcal{F}_k) \geqslant \sigma > 0$, and $\limsup_{k\to\infty} E(\|N_{k+1}\|^\alpha \mid \mathcal{F}_k) < \infty$ for some $\alpha > 2$. Then,* A3 *holds.*

Proof. Let x_k and y_k be $\mathcal{F}_k$-measurable and satisfy (81). By the Hölder inequality with $p = \frac{\alpha-1}{\alpha-2}$ and $q = \alpha - 1$ we have

$$E(|x_k^T N_{k+1} y_k|^2 \mid \mathcal{F}_k) \leqslant (E(|x_k^T N_{k+1} y_k| \mid \mathcal{F}_k))^{\frac{\alpha-2}{\alpha-1}} (E(|x_k^T N_{k+1} y_k|^\alpha \mid \mathcal{F}_k))^{\frac{1}{\alpha-1}}.$$

Since $\limsup_{k\to\infty} E(\|N_{k+1}\|^\alpha \mid \mathcal{F}_k) < \infty$, by (81) for the proposition it suffices to show

$$\liminf_{k\to\infty} E(|x_k^T N_{k+1} y_k|^2 \mid \mathcal{F}_k) > 0.$$

Writing $x_k = [x_1(k), \dots, x_n(k)]^T$ and $y_k = [y_1(k), \dots, y_n(k)]^T$ and noticing the conditional uncorrelatedness of components of N_k, we have

$$\begin{aligned}
\liminf_{k\to\infty} E(|x_k^T N_{k+1} y_k|^2 \mid \mathcal{F}_k) &= \liminf_{k\to\infty} E\left(\sum_{i,j}^n x_i(k) y_j(k) N_{ij}(k+1) \sum_{s,t}^n x_s(k) y_t(k) N_{st}(k+1) \;\Big|\; \mathcal{F}_k \right) \\
&= \liminf_{k\to\infty} \sum_{ij}^n (x_i(k) y_j(k))^2 E(N_{ij}^2(k+1) \mid \mathcal{F}_k) \geqslant \liminf_{k\to\infty} \sum_{i=1}^n x_i^2(k) \sum_{j=1}^n y_j^2(k) \sigma^2 \\
&= \liminf_{k\to\infty} \|x_k\|^2 \|y_k\|^2 \sigma^2 > 0.
\end{aligned}$$

This proves the proposition. $\square$

Prior to describing the result on ordered convergence let us diagonalize the matrix

$$A = [\phi_1, \dots, \phi_n] \begin{bmatrix} \lambda^{(1)} & & & \\ & \lambda^{(2)} & & \\ & & \ddots & \\ & & & \lambda^{(n)} \end{bmatrix} \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_n^T \end{bmatrix}$$

where ϕ_i is a unit eigenvector corresponding to the eigenvalue $\lambda^{(i)}$.

For any j if there is an i such that $\lambda^{(i)} > \lambda^{(j)}$, then we define

$$a(j) = \max\{i\colon \lambda^{(i)} > \lambda^{(j)}\}.$$

Similarly, define

$$b(j) = \min\{i\colon \lambda^{(i)} < \lambda^{(j)}\},$$

if there is an i such that $\lambda^{(i)} < \lambda^{(j)}$.

Let S^i be the set of unit eigenvectors corresponding to $\lambda^{(i)}$, where the identity among some of $\{S^i\}$ is not excluded. Further, let J be the totality of all unit eigenvectors of A. Then

$$J = S^1 \cup \cdots \cup S^n.$$

Theorem 2. *Assume* A1–A3 *hold. Then* $\lambda(i) = \lambda^{(i)}$, $J_i = S^i$, *and* $d(u_k^{(i)}, S^i) \xrightarrow[k\to\infty]{} 0$.

Proof. If $\lambda^{(1)} = \lambda^{(2)} = \cdots = \lambda^{(n)}$, then the conclusion of the theorem follows from Theorem 1. So, we need only to consider the case where $b(1)$ is well defined. Let us first prove the theorem for $i = 1$.

By Theorem 1 we have $d(u_k^{(1)}, J_1) \xrightarrow[k\to\infty]{} 0$. To prove $J_1 = S^1$, it suffices to show $P(d(u_k^{(1)}, S^m) \to 0) = 0$ for any $m \geqslant b(1)$.

Assume the converse: there is an $m \geqslant b(1)$ such that $d(u_k^{(1)}, S^m) \to 0$ a.s. This means that $\phi^T u_k^{(1)} \to 0$ for any $\phi \in S^1$.

Multiplying ϕ^T to both sides of (9) from the left and noticing $A\phi = \lambda^{(1)}\phi$ we derive

$$\begin{aligned}\phi^T u_{k+1}^{(1)} &= \phi^T u_k^{(1)} + a_k\phi^T\big(Au_k^{(1)} - \big(u_k^{(1)T}Au_k^{(1)}\big)u_k^{(1)}\big) + a_k\phi^T\big(\varepsilon_{k+1}^{(1)} + O(a_k)\big)\\ &= \phi^T u_k^{(1)} + a_k\big[\lambda^{(1)} - \big(u_k^{(1)T}Au_k^{(1)}\big)\big]\phi^T u_k^{(1)} + a_k\phi^T\big(\varepsilon_{k+1}^{(1)} + O(a_k)\big).\end{aligned} \tag{82}$$

Set $\Gamma_1 \triangleq \{\omega\colon \phi^T u_k^{(1)} \to 0\}$.

By definition of $\varepsilon_{n+1}^{(1)}$ given after (9) it is clear that

$$\begin{aligned}E\big(\big\|\phi^T\varepsilon_{k+1}^{(1)}\big\| \bigm| \mathcal{F}_k\big) &= E\big(\big\|\phi^T N_{k+1}u_k^{(1)} - \big(u_k^{(1)T}N_{k+1}u_k^{(1)}\big)\phi^T u_k^{(1)}\big\| \bigm| \mathcal{F}_k\big)\\ &\geqslant E\big(\big\|\phi^T N_{k+1}u_k^{(1)}\big\| \bigm| \mathcal{F}_k\big) - E\big(\big|u_k^{(1)T}N_{k+1}u_k^{(1)}\big| \bigm| \mathcal{F}_k\big)\big\|\phi^T u_k^{(1)}\big\|.\end{aligned}$$

Thus, on Γ_1 by A3 we have $\liminf_n E\{\|\varepsilon_{n+1}^{(1)T}\phi\| \mid \mathcal{F}_n\} > 0$.

Further, by A2 $\{\phi^T\varepsilon_{n+1}^{(1)}\}$ is an mds with $E(\|\phi^T\varepsilon_{n+1}^{(1)}\|^2 \mid \mathcal{F}_n) < \infty$ and $\lambda^{(1)} - (u_k^{(1)T}Au_k^{(1)}) \to \lambda^{(1)} - \lambda^{(m)} > 0$.

Then, $P(\Gamma_1) = 0$ by Lemma 1. This means that with probability one $\phi^T u_k^{(1)}$ does not converge to zero. The obtained contradiction shows that $u^{(1)} \in S^1$ a.s.

Inductively assume

$$u^{(i)} \in S^i \quad \text{a.s.}\ \forall i = 1, \ldots, j,\ j \geqslant 1.$$

We now show it holds also for $i = j+1$.

We distinguish two cases: (1) $b(j+1)$ is not defined, and (2) $b(j+1)$ is well defined.

(1) Consider the case where $b(j+1)$ is not defined. Since not all eigenvalues are equal, $a(j+1) \leqslant j$ must be well defined and

$$\lambda^{(1)} \geqslant \cdots \geqslant \lambda^{(a(j+1))} > \lambda^{(a(j+1)+1)} = \cdots = \lambda^{(j+1)} = \cdots = \lambda^{(n)}.$$

The subspace of unit eigenvectors $\{S^1 \cup \cdots \cup S^{a(j+1)}\}$ corresponding to the eigenvalues $\{\lambda^{(1)}, \ldots, \lambda^{(a(j+1))}\}$ is of dimension $a(j+1)$. By (18) the unit vectors $\{u_k^{(1)}, \ldots, u_k^{(a(j+1))}\}$ are asymptotically orthogonal and by the inductive assumptions they converge to $\{S^1 \cup \cdots \cup S^{a(j+1)}\}$. By Theorem 1 $u_k^{(j+1)}$ converges to J_{j+1}. If $d(u_k^{(j+1)}, \{S^1 \cup \cdots \cup S^{a(j+1)}\}) \xrightarrow[k\to\infty]{} 0$, then the set of $a(j+1)+1$ asymptotically orthogonal unit vectors $\{u_k^{(1)}, \ldots, u_k^{(a(j+1))}, u_k^{(j+1)}\}$ would converge to a subspace of dimension $a(j+1)$. This is impossible.

Therefore, $J_{j+1} \subset \{S^{a(j+1)+1} \cup \cdots \cup S^n\}$. Since $\lambda^{(a(j+1)+1)} = \cdots = \lambda^{(j+1)} = \cdots = \lambda^{(n)}$, then $J_{j+1} = S^{j+1}$ and $d(u_k^{(j+1)}, S^{j+1}) \xrightarrow[k\to\infty]{} 0$. In this case the induction is completed.

(2) We now complete the induction by considering the case where $b(j+1)$ is well defined.

By Theorem 1, $d(u_k^{(j+1)}, J_{j+1}) \xrightarrow[k\to\infty]{} 0$ a.s. Since $u_k^{(j+1)T}u_k^{(i)} \xrightarrow[k\to\infty]{} 0$, $\forall i = 1, \ldots, j$ and by the inductive assumption $d(u_k^{(i)}, S^i) \xrightarrow[k\to\infty]{} 0$, $\forall i = 1, \ldots, j$, the converse assumption $d(u_k^{(j+1)}, S^{j+1}) \not\xrightarrow[k\to\infty]{} 0$ is equivalent to $d(u_k^{(j+1)}, S^m) \xrightarrow[k\to\infty]{} 0$ for some $m \geqslant b(j+1)$. This in turn is equivalent to $\phi^T u_k^{(j+1)} \to 0$ for those $\phi \in S^{j+1}$ for which $\phi^T u_k^{(i)} \to 0$, $\forall i = 1, \ldots, j$.

By using the expression of $u_{k+1}^{(j+1)}$ given by the second equality of (45) it follows that

$$\begin{aligned}
P_{k+1}^{(j)} u_{k+1}^{(j+1)} &= P_{k+1}^{(j)}\big[P_k^{(j)} u_k^{(j+1)} + a_k P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} - a_k\big(u_k^{(j+1)T} P_k^{(j)} A_{k+1} P_k^{(j)} \cdot u_k^{(j+1)}\big) P_k^{(j)} u_k^{(j+1)} + O\big(a_k^2\big)\big] \\
&= P_k^{(j)} u_k^{(j+1)} + a_k A P_k^{(j)} u_k^{(j+1)} - a_k\big(u_k^{(j+1)T} P_k^{(j)} A P_k^{(j)} u_k^{(j+1)}\big) P_k^{(j)} u_k^{(j+1)} + a_k \varepsilon_{k+1}^{(j+1)''} \\
&\quad - a_k V_k^{(j)} V_k^{(j)+} A P_k^{(j)} u_k^{(j+1)} - V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)} - a_k V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} \\
&\quad + a_k\big(u_k^{(j+1)T} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}\big) \cdot V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)} + O\big(a_k^2\big),
\end{aligned} \tag{83}$$

where $\varepsilon_{k+1}^{(j+1)''} = P_k^{(j)} N_{k+1} P_k^{(j)} u_k^{(j+1)} - (u_k^{(j+1)T} P_k^{(j)} N_{k+1} P_k^{(j)} u_k^{(j+1)}) P_k^{(j)} u_k^{(j+1)}$ is an mds measurable with respect to $\mathcal{F}_{k+1}$.

Multiplying both sides of (83) by ϕ^T from the left and noticing $A\phi = \lambda^{(j+1)}\phi$ we derive

$$\begin{aligned}
\phi^T P_{k+1}^{(j)} u_{k+1}^{(j+1)} &= \phi^T P_k^{(j)} u_k^{(j+1)} + a_k \phi^T A P_k^{(j)} u_k^{(j+1)} - a_k\big(u_k^{(j+1)T} P_k^{(j)} A P_k^{(j)} u_k^{(j+1)}\big) \\
&\quad \cdot \phi^T P_k^{(j)} u_k^{(j+1)} + a_k \phi^T \varepsilon_{k+1}^{(j+1)''} - a_k \phi^T V_k^{(j)} V_k^{(j)+} A P_k^{(j)} u_k^{(j+1)} - \phi^T \\
&\quad \cdot V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)} - a_k \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} \\
&\quad + a_k\big(u_k^{(j+1)T} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}\big) \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)} + O\big(a_k^2\big) \\
&= \phi^T P_k^{(j)} u_k^{(j+1)} + a_k\big[\lambda^{(j+1)} - \big(u_k^{(j+1)T} P_k^{(j)} A P_k^{(j)} u_k^{(j+1)}\big)\big] \phi^T P_k^{(j)} u_k^{(j+1)} \\
&\quad - a_k \phi^T V_k^{(j)} V_k^{(j)+} A P_k^{(j)} P_k^{(j)} u_k^{(j+1)} - \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} P_k^{(j)} u_k^{(j+1)} \\
&\quad + a_k \phi^T \varepsilon_{k+1}^{(j+1)''} + O\big(a_k^2\big) - a_k \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} \\
&\quad + a_k\big(u_k^{(j+1)T} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}\big) \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)}.
\end{aligned} \tag{84}$$

By (64) (or its equivalent expression (62)) it is known that $a_k \phi^T V_k^{(j)} V_k^{(j)+} A P_k^{(j)} P_k^{(j)} u_k^{(j+1)} = o(a_k) \phi^T P_k^{(j)} u_k^{(j+1)}$, and by (65) (or its equivalent form (63)) $-a_k \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)} + a_k (u_k^{(j+1)T} P_k^{(j)} A_{k+1} P_k^{(j)} u_k^{(j+1)}) \phi^T V_{k+1}^{(j)} V_{k+1}^{(j)+} P_k^{(j)} u_k^{(j+1)} = O(a_k^2)$, we can rewrite (84) as

$$\begin{aligned}
\phi^T P_{k+1}^{(j)} u_{k+1}^{(j+1)} &= \phi^T P_k^{(j)} u_k^{(j+1)} + a_k\big[\lambda^{(j+1)} - u_k^{(j+1)T} P_k^{(j)} A P_k^{(j)} u_k^{(j+1)} + o(1)\big] \\
&\quad \cdot \phi^T P_k^{(j)} u_k^{(j+1)} + O\big(a_k^2\big) + a_k \phi^T \varepsilon_{k+1}^{(j+1)},
\end{aligned} \tag{85}$$

where $\varepsilon_{k+1}^{(j+1)} = \varepsilon_{k+1}^{(j+1)''} - \varepsilon_{k+1}^{(j+1)'} P_k^{(j)} u_k^{(j+1)}$ is a linear combination of mds' measurable with respect to $\mathcal{F}_{k+1}, \mathcal{F}_k, \ldots, \mathcal{F}_{k+2-j}$.

Set $\Gamma_{j+1} = \{\omega\colon\ \phi^T u_k^{(j+1)} \to 0\}$.

We have

$$\phi^T P_k^{(j)} u_k^{(j+1)} = \phi^T u_k^{(j+1)} - \phi^T V_k^{(j)} V_k^{(j)+} u_k^{(j+1)} \to 0 \quad \text{on } \Gamma_{j+1}, \tag{86}$$

because $\phi^T u_k^{(j+1)} \to 0$ on Γ_{j+1} and $V_k^{(j)} V_k^{(j)+} u_k^{(j+1)} \to 0$ by (20).

Notice that

$$\begin{aligned}
E\big(\big\|\phi^T \varepsilon_{k+1}^{(j+1)}\big\| \bigm| \mathcal{F}_k\big) &\geqslant E\big(\big\|\phi^T \varepsilon_{k+1}^{(j+1)''}\big\| \bigm| \mathcal{F}_k\big) - E\big(\big\|\phi^T \varepsilon_{k+1}^{(j+1)'}\big\| \bigm| \mathcal{F}_k\big) \\
&\geqslant E\big(\big\|\phi^T P_k^{(j)} N_{k+1} P_k^{(j)} u_k^{(j+1)}\big\| \bigm| \mathcal{F}_k\big) - E\big(\big|u_k^{(j+1)T} P_k^{(j)} N_{k+1} P_k^{(j)} u_k^{(j+1)}\big| \bigm| \mathcal{F}_k\big) \big|\phi^T P_k^{(j)} u_k^{(j+1)}\big| \\
&\quad - E\big(\big\|\phi^T \varepsilon_{k+1}^{(j+1)'}\big\| \bigm| \mathcal{F}_k\big).
\end{aligned} \tag{87}$$

By Lemma 2 each term in $\varepsilon_{k+1}^{(j+1)'}$ is headed by one of the vectors $\{u_{k-j}^{(1)}, \ldots, u_{k-j}^{(j)}\}$. Since $\phi^T u_k^{(i)} \to 0$, $\forall i = 1, \ldots, j$, it follows that $\|\phi^T \varepsilon_{k+1}^{(j+1)'}\| \xrightarrow[k\to\infty]{} 0$, and by the dominated convergence theorem $E(\|\phi^T \varepsilon_{k+1}^{(j+1)'}\| \mid \mathcal{F}_k) \xrightarrow[k\to\infty]{} 0$.

Then by (86) from (87) it follows that on Γ_{j+1}

$$\liminf_{k\to\infty} E\big(\big\|\phi^T \varepsilon_{k+1}^{(j+1)}\big\| \bigm| \mathcal{F}_k\big) \geqslant \liminf_{k\to\infty} E\big(\big\|\phi^T P_k^{(j)} N_{k+1} P_k^{(j)} u_k^{(j+1)}\big\| \bigm| \mathcal{F}_k\big). \tag{88}$$

Since $\phi^T u_k^{(i)} \to 0$, $\forall i = 1, \ldots, j$, by induction it is directly verified that $\phi^T V_k^{(i)} \xrightarrow[k\to\infty]{} 0$, which implies $\phi^T P_k^{(j)} - \phi^T \xrightarrow[k\to\infty]{} 0$. Further, by (20) it is seen that $P_k^{(j)} u_k^{(j+1)} - u_k^{(j+1)} \xrightarrow[k\to\infty]{} 0$.

Then by A3 from (88) we find that $\liminf_{k\to\infty} E(\|\phi^T \varepsilon_{k+1}^{(j+1)}\| \mid \mathcal{F}_k) > 0$.

By A2 $\limsup_n E(\|\phi^T \varepsilon_{n+1}^{(j+1)}\|^2 \mid \mathcal{F}_n) < \infty$, and by noticing $P_k^{(j)} u_k^{(j+1)} - u_k^{(j+1)} \xrightarrow[k\to\infty]{} 0$, we have $\lambda^{(j+1)} - u_k^{(j+1)T} P_k^{(j)} A P_k^{(j)} u_k^{(j+1)} + o(1) \to \lambda^{(j+1)} - \lambda^{(m)} > 0$. Then, by Lemma 1 we conclude that $P(\Gamma_{j+1}) = 0$, i.e., $\phi^T u_k^{(j+1)}$ cannot converge to zero. The obtained contradiction shows that $u^{(j+1)} \in S^{j+1}$ a.s. □

5. Convergence rates

From the viewpoint of PCA the case $\{\lambda^{(1)} = \cdots = \lambda^{(n)}\}$ is less interesting, because this case means that all components are equally important and PCA may play no roll. Except this less interesting case, we now show that $u_k^{(i)}$ converges to some $u^{(i)} \in S^i$ as k tends to infinity with the rate of convergence pointed out.

Lemma 3. *Let $\{x_k\}$ be recursively defined by*

$$x_{k+1} = x_k + O\left(a_k^{1+\delta}\right) + a_k w_k \quad \text{with } \delta \in \left(0, \frac{1}{2}\right), \tag{89}$$

where $\{w_k\}$ is such that $W_n \triangleq \sum_{k=1}^n a_k^{1-\delta} w_k$ converges to a finite limit as n tends to ∞, and $\{a_k\}$ is given by A1. *Then x_k converges to a vector x and $\|x_k - x\| = O(a_k^\delta)$.*

Proof. Summing up both sides of (89) from 1 to n leads to

$$x_{n+1} = x_1 + O\left(\sum_{k=1}^n a_k^{1+\delta}\right) + \sum_{k=1}^n a_k w_k. \tag{90}$$

By A1 there are a small enough $\epsilon > 0$ and a sufficiently large N so that $a_{i+1}^{-1} - a_i^{-1} > a - \epsilon$, $\forall i \geqslant N$. Then for $n \geqslant N+1$ we have

$$\begin{aligned} a_n^{-\delta} \sum_{k=n}^\infty a_k^{1+\delta} &< \frac{1}{a-\epsilon} a_n^{-\delta} \sum_{i=n}^\infty \left(\frac{1}{a_i^{-1}}\right)^{1+\delta} \left(a_i^{-1} - a_{i-1}^{-1}\right) \\ &\leqslant \frac{1}{a-\epsilon} a_n^{-\delta} \sum_{i=n}^\infty \int_{a_{i-1}^{-1}}^{a_i^{-1}} \left(\frac{1}{a_i^{-1}}\right)^{1+\delta} dx \\ &\leqslant \frac{1}{a-\epsilon} a_n^{-\delta} \int_{{a_{n-1}}^{-1}}^\infty \left(\frac{1}{x}\right)^{1+\delta} dx = \frac{1}{\delta(a-\epsilon)} \left(\frac{a_{n-1}}{a_n}\right)^\delta = O(1). \end{aligned} \tag{91}$$

Summing by parts, we have

$$\begin{aligned} a_n^{-\delta} \sum_{k=n}^\infty a_k w_k &= a_n^{-\delta} \sum_{k=n}^\infty (W_k - W_{k-1}) a_k^\delta \\ &= a_n^{-\delta} \sum_{k=n}^\infty W_k\left(a_k^\delta - a_{k+1}^\delta\right) - W_{n-1} \\ &= a_n^{-\delta} \sum_{k=n}^\infty W_k a_k^\delta \left(1 - \left(\frac{a_{k+1}}{a_k}\right)^\delta\right) - W_{n-1} = O(1), \end{aligned} \tag{92}$$

because $a_{i+1}^{-1} - a_i^{-1} \xrightarrow[k\to\infty]{} a > 0$ and from here it follows that $1 - (\frac{a_{k+1}}{a_k})^\delta - a\delta a_{k+1} = o(a_k)$.

From (92) it is seen that $\sum_{k=1}^\infty a_k w_k$ converges and the rate of convergence is $\sum_{k=n}^\infty a_k w_k = O(a_n^\delta)$. Then, from (90) we conclude that x_k converges to a vector denoted by x, and

$$x - x_n = O\left(\sum_{k=n}^\infty a_k^{1+\delta}\right) + \sum_{k=n}^\infty a_k w_k. \tag{93}$$

Thus, by (91) and (92) we derive $\|x_k - x\| = O(a_k^\delta)$. □

Theorem 3. *Assume* A1–A3 *hold. Except the case $\{\lambda^{(1)} = \cdots = \lambda^{(n)}\}$, there are $u^{(i)} \in S^i$, $i = 1, \ldots, n$ such that*

$$\lim_{k\to\infty} \left\| u_k^{(i)} - u^{(i)} \right\| = O\left(a_k^\delta\right) \quad \text{with some } \delta \in \left(0, \frac{1}{2}\right),\ \forall i = 1, \ldots, n.$$

Proof. Since the case $\{\lambda^{(1)} = \cdots = \lambda^{(n)}\}$ is excluded from consideration, $b(1)$ is well defined.

Multiplying both sides of (9) from left by $\phi_{b(1)}^T$ and noticing $A\phi_{b(1)} = \lambda^{b(1)}\phi_{b(1)}$ we derive

$$\begin{aligned}
\phi_{b(1)}^T u_{k+1}^{(1)} &= \phi_{b(1)}^T\big[u_k^{(1)} + a_k\big(Au_k^{(1)} - \big(u_k^{(1)T}Au_k^{(1)}\big)u_k^{(1)}\big) + a_k\varepsilon_{k+1}^{(1)} + O\big(a_k^2\big)\big] \\
&= \phi_{b(1)}^T u_k^{(1)} + a_k\big[\lambda^{b(1)} - \big(u_k^{(1)T}Au_k^{(1)}\big)\big]\big(\phi_{b(1)}^T u_k^{(1)}\big) + a_k\phi_{b(1)}^T\varepsilon_{k+1}^{(1)} + O\big(a_k^2\big) \\
&= \phi_{b(1)}^T u_k^{(1)} + a_k\big[\big(\lambda^{b(1)} - \lambda^{(1)}\big)\big(\phi_{b(1)}^T u_k^{(1)}\big) + \big(\lambda^{(1)} - u_k^{(1)T}Au_k^{(1)}\big)\cdot\big(\phi_{b(1)}^T u_k^{(1)}\big)\big] \\
&\quad + a_k\phi_{b(1)}^T\varepsilon_{k+1}^{(1)} + O\big(a_k^2\big),
\end{aligned} \tag{94}$$

where $\lambda^{b(1)} - \lambda^{(1)} < 0$ and $\lambda^{(1)} - u_k^{(1)T}Au_k^{(1)} \xrightarrow[k\to\infty]{} 0$.

Define $\epsilon'_{k+1} \triangleq \phi_{b(1)}^T\varepsilon_{k+1}^{(1)}$, $\epsilon''_{k+1} = O(a_k)$.

By A2, ϵ'_{k+1} is an mds measurable with respect to $\mathcal{F}_{k+1}$.

Since $\sum_k a_k^{2(1-\delta)} < \infty$ for any $\delta \in (0, \frac{1}{2})$, $\sup_k E\{\|N_{k+1}\|^2 \mid \mathcal{F}_k\} < \infty$, and $\|u_k^{(1)}\| = 1$, by the convergence theorem for mds' we have $\sum_k a_k^{1-\delta}\epsilon'_{k+1} < \infty$.

Further, $\epsilon''_{k+1} = O(a_k) = o(a_k^\delta)$, and by Theorem 3.1.1 in [3] we find $\phi_{b(1)}^T u_k^{(1)} = o(a_{k-1}^\delta)$.

Similarly, we obtain $\phi_i^T u_k^{(1)} = O(a_k^\delta)$, $i = b(1)+1, \ldots, n$.

Consequently, we derive

$$d\big(u_k^{(1)}, S^1\big) = \sqrt{\big(\phi_{b(1)}^T u_k^{(1)}\big)^2 + \cdots + \big(\phi_n^T u_k^{(1)}\big)^2} = O\big(a_k^\delta\big).$$

Then, there is a subsequence $u^{(1)}(k) \in S^1$ such that $\|u_k^{(1)} - u^{(1)}(k)\| = O(a_k^\delta)$ and $\|Au_k^{(1)} - (u_k^{(1)T}Au_k^{(1)})u_k^{(1)}\| = O(\|u_k^{(1)} - u^{(1)}(k)\|) = O(a_k^\delta)$.

From here and (8) it follows that

$$u_{k+1}^{(1)} = u_k^{(1)} + O\big(a_k^{1+\delta}\big) + a_k\delta_{k+1}(1).$$

Since $\sum_{k=1}^\infty a_k^{1-\delta}\delta_{k+1}(1) < \infty$, by Lemma 3 the theorem is true for $u_k^{(1)}$.

Inductively assume that

$$\big\|u_k^{(i)} - u^{(i)}\big\| = O\big(a_k^\delta\big), \quad \forall i = 1, \ldots, j$$

for some $\delta > 0$. We now show that it also holds for $i = j+1$.

By (38) it is seen that

$$u^{(i)T}u_{k+1}^{(j+1)} = \big(u^{(i)} - u_k^{(i)} + u_k^{(i)}\big)^T u_{k+1}^{(j+1)} = O\big\|u_k^{(i)} - u^{(i)}\big\| = O\big(a_k^\delta\big), \quad \forall i = 1, \ldots, j. \tag{95}$$

(1) If $b(j+1)$ is not defined, then $a(j+1) \leqslant j$ must be well defined. In this case,

$$d^2\big(u_k^{(j+1)}, S^{j+1}\big) = O\left(\sum_{i=1}^{a(j+1)}\big(u^{(i)T}u_{k+1}^{(j+1)}\big)^2\right) = O\big(a_k^{2\delta}\big), \quad \delta > 0, \tag{96}$$

where (95) and the inductive assumption are used.

Similar to $u_k^{(1)}$, there is a sequence $u^{(j+1)}(k) \in S^{j+1}$ such that $\|u_k^{(j+1)} - u^{(j+1)}(k)\| = O(a_k^\delta)$, and hence $\|Au_k^{(j+1)} - (u_k^{(j+1)T}Au_k^{(j+1)})u_k^{(j+1)}\| = O(\|u_k^{(j+1)} - u^{(j+1)}(k)\|) = O(a_k^\delta)$, and by (8)

$$u_{k+1}^{(j+1)} = u_k^{(j+1)} + O\big(a_k^{1+\delta}\big) + a_k\delta_{k+1}(j+1).$$

Then, again by Lemma 3 the conclusion of the theorem follows for $j+1$. The induction is completed for this case.

(2) Consider the case where $b(j+1)$ is well defined.

Let ϕ be any unit vector from $S^{b(j+1)+i}$, $\forall i = 0, 1, \ldots, n - b(j+1)$. By the inductive assumption from (8) it is seen that

$$u_{k+1}^{(j+1)} = u_k^{(j+1)} + a_k\big(Au_k^{(j+1)} - \big(u_k^{(j+1)T}Au_k^{(j+1)}\big)u_k^{(j+1)}\big) + O\big(a_k^{1+\delta}\big) + a_k\delta_{k+1}(j+1). \tag{97}$$

Proceeding as in (94) we derive

$$\begin{aligned}
&\phi_{b(j+1)+i}^T u_{k+1}^{(j+1)} \\
&\quad = \phi_{b(j+1)+i}^T u_k^{(j+1)} + a_k\big[\lambda^{b(j+1)+i} - \big(u_k^{(j+1)T} A u_k^{(j+1)}\big)\big]\big(\phi_{b(j+1)+i}^T u_k^{(j+1)}\big) + a_k \phi_{b(j+1)+i}^T \delta_{k+1}(j+1) + O\big(a_k^{1+\delta}\big) \\
&\quad = \phi_{b(j+1)+i}^T u_k^{(j+1)} + a_k\big[\big(\lambda^{b(j+1)+i} - \lambda^{(j+1)}\big)\big(\phi_{b(j+1)+i}^T u_k^{(j+1)}\big) + \big(\lambda^{(j+1)} - u_k^{(j+1)T} A u_k^{(j+1)}\big)\big(\phi_{b(j+1)+i}^T u_k^{(j+1)}\big)\big] \\
&\qquad + a_k \phi_{b(j+1)+i}^T \delta_{k+1}(j+1) + O\big(a_k^{1+\delta}\big),
\end{aligned}$$

where $\lambda^{b(j+1)+i} - \lambda^{(j+1)} < 0$ and $\lambda^{(j+1)} - u_k^{(j+1)T} A u_k^{(j+1)} \xrightarrow[k\to\infty]{} 0$.

Again by the convergence rate theorem [3] we have

$$\big|\phi_{b(j+1)+i}^T u_k^{(j+1)}\big| = O\big(a_k^\delta\big). \tag{98}$$

Finally, we conclude that

$$d^2\big(u_k^{(j+1)}, S^{j+1}\big) = O\left(\sum_{i=1}^{a(j+1)} \big(u_k^{(j+1)T} u^{(i)}\big)^2\right) + O\left(\sum_{i=0}^{n-b(j+1)} \big(\big(\phi_{b(j+1)+i}^T u_k^{(j+1)}\big)\big)^2\right) = O\big(a_k^{2\delta}\big), \quad \delta > 0$$

by (95), (98) and the inductive assumption.

This is the same estimate as (96), and from here we find that there is a $u^{(j+1)} \in S^{j+1}$ such that $\|u_k^{(j+1)} - u^{(j+1)}\| = O(a_k^\delta)$. □

6. Concluding remarks

The recursive algorithms based on stochastic approximation are given in the paper to estimate the eigenvectors and eigenvalues of a symmetric matrix A, which is observed with additive noise. The matrix A is allowed to have arbitrary eigenvalues. The a.s. ordered convergence with rate of convergence is established except the special case where all eigenvalues are equal.

Though this special case is less interesting for PCA, from the theoretical point of view it is still of some interest to clarify whether the estimates converge to limits or not. It is also of interest to weaken conditions imposed on N_k, for example, the boundedness of N_k, and the conditions specified in A3.

Appendix A

The proof of Lemma 1 is based on the following fact. We formulate it as Lemma 4, but for its proof we refer to [1,2].

Lemma 4. *Let $\{\mathcal{F}_k\}$ be a family of nondecreasing σ-algebras, and let $\{\varepsilon_k, \mathcal{F}_k\}$ be an mds satisfying*

$$E\big\{\|\varepsilon_{k+1}\|^2 \bigm| \mathcal{F}_k\big\} < \infty \quad \text{and} \quad E\{\varepsilon_{k+1} \mid \mathcal{F}_k\} = 0. \tag{99}$$

Let the real sequence $\{a_k\}$ be such that $a_k > 0$, $\sum_k a_k = \infty$, $\sum_k a_k^2 < \infty$, and let $\{\Theta_k, \mathcal{F}_k\}$ be an adapted sequence.

Assume that on $\Gamma \subset \Omega$ the following conditions are satisfied.

(1) $\limsup_{k\to\infty} E\{\|\varepsilon_{k+1}\|^2 \mid \mathcal{F}_k\} < \infty$, $\liminf_{k\to\infty} E\{\|\varepsilon_{k+1}\| \mid \mathcal{F}_k\} > 0$.
(2) *For some n, $\sum_{k=n}^\infty a_k(\Theta_k + \varepsilon_k)$ coincides with an $\mathcal{F}_n$-measurable random variable.*
(3) *Θ_k can be decomposed into two adapted sequences $\{r_k, \mathcal{F}_k\}$ and $\{R_k, \mathcal{F}_k\}$: $\Theta_k = r_k + R_k$ such that*

$$\sum_k \|r_k\|^2 < \infty \quad \text{and} \quad E\left\{I_\Gamma \sum_{k=n}^\infty \|a_k R_k\|\right\} = o\left(\sum_{k=n}^\infty a_k^2\right)^{\frac{1}{2}} \quad \text{as } n \to \infty.$$

Then, $P(\Gamma) = 0$.

Proof of Lemma 1. Let $\Gamma = \{\omega\colon\ X_k \to 0\}$ and $\Phi(n+1, n_0) = \prod_{k=n_0}^n (1 + a_k\alpha_k)$ with $\Phi(n_0, n_0) = 1$. Then we have

$$X_n = \Phi(n, n_0)\left[X_{n_0} + \sum_{j=n_0}^{n-1} \Phi^{-1}(j+1, n_0)\big(a_j\varepsilon_{j+1} + O\big(a_j^2\big)\big)\right]. \tag{100}$$

Noticing that $X_n \to 0$ at Γ and $\Phi(n+1, n_0) \to \infty$ as $n \to \infty$, from (100) we obtain

$$X_{n_0} = -\sum_{j=n_0}^\infty \Phi^{-1}(j+1, n_0)\big(a_j\varepsilon_{j+1} + O\big(a_j^2\big)\big), \quad \forall n_0 \geqslant 0.$$

Rewriting n_0 as n, and setting $\Phi_j \triangleq \Phi(j+1,n)$ and $S_j = \sum_{k=j}^{\infty}(a_k\varepsilon_{k+1} + O(a_k^2))$, we have

$$X_n = \sum_{j=n}^{\infty}\Phi_j^{-1}[S_{j+1} - S_j] = -S_n + \sum_{j=n}^{\infty}\big(\Phi_j^{-1} - \Phi_{j+1}^{-1}\big)S_{j+1}. \tag{101}$$

Set $\varepsilon'_{k+1} = \sum_{i=1}^{l}\frac{a_{k+i-1}}{a_k}\varepsilon_{k+1}^{(i)}$. It is clear that $\{\varepsilon'_k, \mathcal{F}_k\}$ is an mds. Noticing $\varepsilon_k^{(i)} \to 0$, $\forall i \geqslant 2$, we have

$$\limsup_{k\to\infty} E\big(\big\|\varepsilon'_{k+1}\big\|^2 \bigm| \mathcal{F}_k\big) < \infty, \qquad \liminf_{k\to\infty} E\big(\big\|\varepsilon'_{k+1}\big\| \bigm| \mathcal{F}_k\big) > 0, \tag{102}$$

and

$$\begin{aligned} X_n &= \sum_{j=n}^{\infty}\big[\big(\Phi_j^{-1} - \Phi_{j+1}^{-1}\big)S_{j+1} - a_j\varepsilon'_{j+1} + O\big(a_j^2\big)\big] + \sum_{j=n}^{\infty}a_j\big(\varepsilon'_{j+1} - \varepsilon_{j+1}\big) \\ &= \sum_{j=n}^{\infty}\big[\big(\Phi_j^{-1} - \Phi_{j+1}^{-1}\big)S_{j+1} - a_j\varepsilon'_{j+1} + O\big(a_j^2\big)\big] - \sum_{s=2}^{l}\sum_{i=0}^{s-2}a_{n+i}\varepsilon_{n-s+i+2}^{(s)}. \end{aligned} \tag{103}$$

Decompose $\Theta_k = R_k + r_k$, where $R_k = \frac{1}{a_k}\{(\Phi_k^{-1} - \Phi_{k+1}^{-1})S_{k+1}\}$ and $r_k = O(a_k)$. Then from (103) we have

$$\sum_{j=n}^{\infty}a_j\big(\Theta_j - \varepsilon'_{j+1}\big) = X_n + \sum_{s=2}^{l}\sum_{i=0}^{s-2}a_{n+i}\varepsilon_{n-s+i+2}^{(s)} \in \mathcal{F}_n. \tag{104}$$

Since $\sum_k \|r_k\|^2 < \infty$, by Lemma 3 we conclude $P(\Gamma) = 0$ if we can show that as $n \leqslant n_1 \to \infty$

$$E\left\{I_\Gamma \sum_{j=n_1}^{\infty}\|a_jR_j\|\right\} = o\left(\sum_{j=n_1}^{\infty}a_k^2\right)^{\frac{1}{2}}. \tag{105}$$

We now prove (105).

Noticing $\Phi_j = \prod_{k=n}^{j}(1 + a_k\alpha_k)$ and $\frac{3}{2}\alpha > \alpha_k > \frac{1}{2}\alpha$, $\forall k \geqslant n$ for sufficiently large n since $\alpha_k \to \alpha > 0$, we then have

$$\begin{aligned} \ln\Phi_j &= \sum_{k=n}^{j}\ln(1 + a_k\alpha_k) = \sum_{k=n}^{j}a_k\alpha_k + O\left(\sum_{k=n}^{j}a_k^2\right) \\ &> \frac{\alpha}{4}\sum_{k=n}^{j}a_k, \end{aligned}$$

and hence $\Phi_j > \exp(\frac{\alpha}{4}\sum_{k=n}^{j}a_k)$ or $\Phi_j^{-1} < \exp(-\frac{\alpha}{4}\sum_{k=n}^{j}a_k)$. Further, by the Schwarz inequality we have

$$\begin{aligned} E\left\{I_\Gamma \sum_{j=n_1}^{\infty}\|a_jR_j\|\right\} &\leqslant \sum_{j=n_1}^{\infty}\big(\big|\Phi_j^{-1} - \Phi_{j+1}^{-1}\big|^2 ES_{j+1}^2\big)^{\frac{1}{2}} \\ &= \sum_{j=n_1}^{\infty}a_{j+1}\alpha_{j+1}\Phi_{j+1}^{-1}\big(ES_{j+1}^2\big)^{\frac{1}{2}} \\ &\leqslant \frac{3}{2}\sum_{j=n_1}^{\infty}\alpha a_{j+1}\exp\left(-\frac{\alpha}{4}\sum_{k=n}^{j+1}a_k\right)\big(ES_{j+1}^2\big)^{\frac{1}{2}}. \end{aligned} \tag{106}$$

By noticing that $\{\varepsilon_k^{(i)}, \mathcal{F}_k\}$ is an mds, for sufficiently large $j \geqslant n_1 \geqslant n$, we have

$$\begin{aligned} ES_{j+1}^2 &\leqslant 2E\left(\sum_{k=j+1}^{\infty}a_k\varepsilon_{k+1}\right)^2 + O\left(\left(\sum_{k=j+1}^{\infty}a_k^2\right)^2\right) \\ &= 2E\left(\sum_{k=j+1}^{\infty}\sum_{i=1}^{l}a_k\varepsilon_{k+2-i}^{(i)}\right)^2 + O\left(\left(\sum_{k=j+1}^{\infty}a_k^2\right)^2\right) \\ &= O\left(\sum_{k=j+1}^{\infty}a_k^2\right). \end{aligned} \tag{107}$$

Since $\lim_{x\to+\infty} x^2\exp(-\frac{1}{4}\alpha x)=0$ and $\sum_k a_k=+\infty$, for large enough $n_1>n$

$$\frac{3}{2}\sum_{j=n_1}^{\infty}\alpha a_{j+1}\exp\left(-\frac{\alpha}{4}\sum_{k=n}^{j+1}a_k\right)\leqslant\frac{3}{2}\alpha\sum_{j=n_1}^{\infty}\frac{a_{j+1}}{(\sum_{k=n}^{j+1}a_k)^2}=\frac{3}{2}\alpha\sum_{j=n_1+1}^{\infty}\int_{\sum_{k=n_1}^{j-1}a_k}^{\sum_{k=n_1}^{j}a_k}\frac{dx}{(\sum_{k=n_1}^{j}a_k)^2}$$
$$\leqslant\frac{3}{2}\alpha\int_{\sum_{k=n}^{n_1+1}a_k}^{\infty}\frac{dx}{x^2}=\frac{3\alpha}{2\sum_{k=n}^{n_1+1}a_k}\xrightarrow[n_1\to\infty]{}0. \tag{108}$$

Combining (106), (107), and (108) leads to (105) and the lemma itself. □

References

[1] O. Brandiere, Some pathological traps for stochastic approximation, SIAM J. Control Optim. 36 (4) (1998) 1293–1314.
[2] O. Brandière, Marie Duflo, Les algorithmes stochastiques contournent-ils les pièges?, Probab. Statist. 32 (1996) 395–427.
[3] H.F. Chen, Stochastic Approximation and Its Applications, Kluwer, Dordrecht, 2002.
[4] A. Daffertshofer, C.J.C. Lamoth, O.G. Meijer, P.J. Beek, PCA in studying coordination and variability: a tutorial, Clin. Biomech. 19 (4) (2004) 415–428.
[5] B. Delyon, General convergence result on stochastic approximation, IEEE Trans. Automat. Control 41 (1996) 1245–1255.
[6] A. Meyer-Baese, Medical image compression by neural-gas network and principal component analysis, in: IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN00), vol. 5, 2000, p. 5489.
[7] E. Oja, Subspace Methods of Pattern Recognition, Research Studies Press and J. Wiley, Letchworth, 1983.
[8] E. Oja, J. Karhunen, On stochastic approximation of the eigenvectors and eigenvalues of the expection of a random matrix, J. Math. Anal. Appl. 106 (1) (1985) 69–84.
[9] K. Pearson, On lines and planes of closest fit to systems of points in space, Philos. Mag. Ser. 6 2 (11) (1901) 559–572.
[10] W.X. Zhao, Identification and adaptive tracking for some classes of stochastic systems, PhD thesis, AMSS, Chinese Academy of Sciences, 2008.
[11] J.H. Zhang, H.F. Chen, Convergence of algorithms used for principal component analysis, Sci. China Ser. E 40 (6) (1997) 597–604.